

Government without land

How will the world cope with the rash of new nations? Israel's tentative scheme for the West Bank, which proposes separate governments and separate taxes in the same land, will not often work.

WHETHER the peace conference that opened in Madrid two weeks ago will settle the Middle East's old scores is far from clear, but some of Israel's early talk of how the West Bank may eventually be administered will be taken seriously in other troubled regions. The notorious difficulty in what used to be Palestine is that both Jewish and Arab people assert equal and conflicting claims to live there. Israel has therefore been canvassing the notion that national sovereignty might be divorced from the ownership of the land. The idea is that both Palestinians and Israelis would continue to live in the disputed territories, but that their interests would be watched over by separately elected governments to which they would presumably separately pay taxes. Only the negotiations that lie ahead will tell whether such a scheme is workable in the region for which it is intended, but it is bound to excite the interest of the clutch of emergent states whose ambitions for independence are compromised by ethnic and cultural disunity.

Yugoslavia is the obvious, but not the only, case. The immediate cause of the conflict in Croatia is given as the presence there of a substantial Serbian minority, but Serbia itself is embarrassed by a substantial minority of mostly Muslim Albanians. Georgia, in the old Soviet Union, has a similar problem, as has Azerbaijan, while the cultural minorities in the Russian Federation proper, many of them arbitrarily contrived (as by Stalin's expulsion of Tartars from the Crimea) will yet come to haunt Mr Boris Yeltsin and his successors. Yet there are other examples of states divided within themselves that are now so familiar that the parallels are easily overlooked. Belgium is now a loose federation of two states divided on language grounds, Britain's Northern Ireland is potentially another, while South Africa's apartheid regime, now mercifully on the way out, was justified as a means of allowing different peoples to coexist.

None of this is encouraging. Although the French-speaking (Walloon) community in Belgium has been for decades less well-off than the Dutch-speaking community, the differences now are not so great that coexistence is a strain. The South African experiment shows that glaring economic disparities (in that case linked to the partition of the land) make stability impossible. In any case, such schemes will function only if sovereignty is constrained. Separate community governments within a single state are usually prevented at their constitution from making common cause with outside governments;

separate Protestant and Catholic governments in Northern Ireland would be politically unacceptable (to the Protestants) if the Catholic government were free to throw in its lot with the Irish Republic to the south, while the economic differences that have grown up with several decades of religious discrimination would not thereby vanish. On the face of things, the Israeli scheme for the West Bank could be made to work only if there were a new state, separate from Israel proper, and a mechanism for redistributing wealth as it is earned between two poor communities, one poorer than the other. Would Israel go that far?

For the rest, the only workable ways of coexistence for interpenetrating distinct communities must be old-fashioned ways. The general alarm about the emergence of small states in Central Europe (west of the Urals) and the Baltic can be dealt with only on the basis that the civil liberties of all who live within a state's boundaries are effectively guaranteed. Luckily, the emerging states are all *de facto* signatories of the Helsinki accords of 1978, which commit them to enlightened policies of civil rights. Is it not time to back these promises with enforcement? And might not the European Communities' attempt to mediate in Yugoslavia have been more successful if means of enforcing fair play were already in force? □

Death of a tycoon

Robert Maxwell's death last week will engender mixed passions in the scientific community, as elsewhere.

Mr Robert Maxwell, the publishing tycoon who died last week, had been an important influence on the business of science since the early 1960s. He was one of the first to recognize that the Soviet Union was bursting with excellent science. His Pergamon Press performed an important public service by translating into English many important Russian works, the influential textbooks of Landau and Lifshitz, for example. Maxwell was also a prolific creator of new scientific journals, some of which (*Tetrahedron*, for example) have excellent reputations. But Maxwell had kleptomaniac tendencies and started or acquired many journals that are undistinguished. And he did not die a scientific publisher; he sold his Pergamon Press to the Netherlands company Elsevier (which has since also acquired the *Lancet*) only a few weeks before his death.

Maxwell will be remembered differently by different



people. To many with outstanding reputations in research, he was a thoroughly admirable person, distinguished by his flair for using power, influence and money to get things done. To others, he was a bully and even a cheat. Among publishers, he was outstanding for his readiness personally to argue the toss about disputed royalty statements with disappointed authors. More recently, he cut the figure of a potential benefactor, dangling beneficence in front of universities (Harvard and Duke for two) and individuals — he kept Dr Robert C. Gallo on a string for several months with a scheme for an independent laboratory that came to nothing.

That Maxwell was indeed a remarkable fellow is not disputed. His energy was prodigious, as was his flair for striking advantageous deals. But he also cut corners shamelessly, as for example when he sought to sell a share in Pergamon Press to the US businessman Saul Steinberg on the basis of a balance-sheet in which unsold books and journals were counted as assets; not all his ebullience quite lived down the complaint of a British government inspector that he was not a "fit person" to run a public company. Shareholders in his companies, and the army of people they employ, will no doubt remember that complaint as they ponder on the struggles of the accountants to tell whether the debts (some £2,000 million) exceed the value of the surviving assets. It is a great pity that one so brave and energetic should so often have been so misguided. □

Burning tritium

Is too much being made of the most recent milestone in fusion development?

NOT for the first time, research into controlled nuclear fusion (the attempt to harness thermonuclear energy such as keeps stars shining) made newspaper headlines. The result, from the Joint European Torus (JET) at Culham in Oxfordshire, England (see page 95) is genuinely important. For 30 years, fusion scientists have shied away from experiments with plasma containing both deuterium and tritium, the fuel of prospective fusion power stations, because of the irradiation of the apparatus that would follow. But now it has been tried, and the results are more than encouraging.

But is there a danger that repeated announcements of new milestones will be counter-productive, enervating rather than enthusing those they are meant to influence? It is only two years since JET announced results also purportedly showing that the reactor was within spitting distance of 'breakeven' — the point at which fusion reactions within the hot plasma produce as much energy as needed to heat the plasma in the first place. But sleight-of-hand was required on that occasion: the researchers would have been within 80 per cent of breakeven had they been using a deuterium-tritium plasma, not one of pure deuterium. Again, the actual achievement was impressive, with the product of the three key parameters in the

quest for fusion power (density, temperature and energy-confinement time) being a record at that time. That result and those just announced apparently bear out the claim to pre-eminence in the field for Europe claimed by Paul Rebut, director of JET. But there is a danger of wearying the public with repeated claims of "nearly there" when the truth is, as anyone will admit, that it will be half a century before fusion reactors are used commercially.

That distant horizon is of course the problem for fusion researchers. Making fusion work means building big machines — either the conventional tokamak design in which the reacting plasma is confined magnetically or the more adventurous laser-driven 'inertial' confinement — and big machines are expensive. The carrot has to be repeatedly held in front of the politicians. In that spirit, the new result could hardly have come at a better time. National representatives are meeting in Moscow this week to decide on funding for the design phase of the next multinational fusion collaboration, the International Thermonuclear Experimental Reactor (ITER). Planned to produce more than 1,000 MW of power — a thousand times the amount just produced at JET — the project could cost \$5,000 million. What better encouragement to give politicians than a new demonstration of the potential of fusion power?

To the researchers, this politicking is tiresome. Paul Rebut wearily admits that decisions on the siting of giant collaborations such as ITER (bringing together Europe, the United States, Japan and the Soviet Union) are not entirely rational. The design phase will be split between Japan, Germany and the United States. And Alan Gibson, associate director of JET, suggests that the cost of fusion research be compared with the turnover of the electricity industry; the ratio, well short of one per cent, would not look out of place in the research-and-development programme of a multinational chemical company.

Even so, researchers in the United States are looking enviously at the new JET results. Equivalent deuterium-tritium plasmas were originally intended for Princeton University's TFTR (Tokamak Fusion Test Reactor) to produce 10 MW in 1986–87, but were turned down by the Department of Energy. Instead, Princeton will have to wait until July 1993 before it starts using tritium, and will take up the last year of that reactor's experimental programme. JET's own large-scale tritium experiments will take place in 1996, also in the last year of operations (extended six years beyond the original deadline of 1990). The present experiments are a one-time only test, timed to allow radioactivity to die away before a substantial refit of the reactor planned for next February.

Beyond that, researchers see a big blank. Princeton's Burning Plasma Experiment (BPX), intended to prepare the ground for ITER, was recently rejected by the US Department of Energy. And construction of ITER will not begin for at least four years. So, unless those working on inertial confinement fusion can pull something out of the hat, there may be a shortage of milestones in the next few years. □

Dingell opens second front in Gallo war

- Focus shifts from misconduct to perjury
- One court has already rejected case

Washington

US CONGRESSIONAL investigators, frustrated by their failure so far to obtain a conclusion of scientific misconduct in the case of AIDS researcher Robert Gallo, have launched a second line of attack: allegations of perjury and patent fraud. Seeking to bring criminal charges against the embattled National Institutes of Health (NIH) scientist, the investigators plan to launch an inquiry into statements Gallo made under oath in documents associated with a 1987 agreement between the United States and France to share credit for the AIDS blood test.

The staff of the congressional oversight and investigations subcommittee of Representative John Dingell (Democrat, Michigan) has asked the congressional General Accounting Office (GAO) to examine a 1986 sworn declaration in which Gallo says that in April 1984 — the date of the US patent application for the AIDS blood test — he believed that the AIDS virus isolated by his own team and that isolated by the French were not “the same, or even substantially the same, virus”.

Both the French virus, which was called LAV at the time, and Gallo’s virus, known then as HTLV-III, have been shown within the past few months to have been derived from the same source — a French patient given the code name LAI, whose aggressive virus apparently contaminated laboratory samples on both sides of the Atlantic, although no one knew it at the time. New studies have shown that the French and US isolates are virtually identical, despite some early evidence to the contrary.

The key question in Dingell’s perjury case is what Gallo knew at the time of the patent application. Although Gallo could not have known that both his and the French virus were actually LAI, the investigators must determine if he nevertheless knew in early 1984 that the two isolates were more alike than the French literature suggested. Did Gallo, for example, choose to cite studies that indicated differences even though he knew that those studies were flawed? One source close to the Dingell investigation claims that evidence now assembled will show that “there were plenty of documents [at the time] to tell Gallo and his people whether the viruses were substantially the same or not.” Neither this source nor others close to the Dingell committee would make these documents available for inspection.

According to his lawyer, Joseph Onek

of the Washington law firm of Crowell and Moring, Gallo based his statement on studies of the French virus published by the team at the Paris-based Institut Pasteur that had isolated it.

In July and September 1983, Luc Montagnier of the Institut Pasteur sent Gallo samples of the virus. Gallo’s colleagues had tested it for reverse transcriptase and examined it under an electron microscope — two procedures that indicated that it was indeed a retrovirus. NIH researchers also tried, during several months, to grow the French virus in a special cell line, known as H4. But by late fall 1983, Gallo had decided that the US isolate, HTLV-III, was a more promising candidate for the basis of an AIDS test, and he ordered that French virus to be removed from study.

Onek says that “there were no comparisons [between the viruses] done [by Gallo’s staff] until May or June 1984 or later. All the data about LAV comes from the French literature.” Those papers, which are referenced in the declaration, reported that only about 20 per cent of AIDS patients had antibodies that reacted with the French isolate (compared to more than 80 per cent reactivity with Gallo’s virus), and that the French virus did not show the presence of gp41, a glycoprotein that Gallo had found in the envelope of his laboratory’s isolate.

Onek says that he cannot rule out the possibility that someone in Gallo’s laboratory or elsewhere could have conceivably done an analysis of the French virus before April 1984 that showed the two viruses to be substantially similar. “It’s possible that someone found gp41 [in the French virus] and didn’t tell anyone. But to this day, we have seen not seen any comparison of the kind that would negate the findings” in the declaration.

Onek acknowledges that there are errors in the declaration but says they are unimportant. He says that some references to the ‘H9’ cell line — one of the most successful lines for growing the AIDS virus — are actually describing an earlier cell line known as ‘H4’. But he argues that this was a mistake that does not significantly change the conclusion of the declaration or the basis of the agreement. Other statements in the document are known to be wrong only because of information that has come to light since 1986, he says. “Are there errors in the declaration? Yes. Were they known at the time? No.”

But in a 100-page document entitled

“The Great AIDS Cover-Up”, members of the Dingell staff have collected evidence they say suggests that Gallo knew more than he admitted. The Dingell report, first described in the *Chicago Tribune* last week, outlines three possible classes of criminal violations, including alleged errors in scientific data contained in a 1984 article in *Science* that formed the basis of the US patent, as well as the possibility that Gallo made false and misleading statements under oath to US officials in the preparation of documents related to the patent application and the agreement to share credit and royalties with the French.

These allegations, however, have already been rejected by one US Attorney. On 28 October, the office of the inspector general for the department of Health and Human Services (HHS), which oversees NIH, presented evidence for a criminal prosecution in this case to the US Attorney for the District of Columbia. According to a letter dated 5 November from HHS general counsel Michael Astrue to Dingell, the US Attorney’s office “decided not to pursue criminal prosecution at this time.”

Dingell’s staff blames this failure to bring criminal charges on a poor presentation by the HHS and OSI officials. But they hope to have more success in a second attempt based on whatever evidence the GAO can assemble. On the advice of the US Department of Justice, the congressional staff will make their new case for a prosecution to the US Attorney in Baltimore, Maryland. This has several advantages, they argue. NIH is physically located in Maryland, and more properly under Maryland jurisdiction than that of the District of Columbia, where HHS is based. And the Maryland US Attorney has had experience with other NIH investigations, including the Thereza Imanishi-Kari scientific misconduct case and two financial irregularity prosecutions involving members of Gallo’s own laboratory.

Christopher Anderson

FUSION

First use of tritium

London

THE first nuclear fusion experiment to use tritium in addition to deuterium produced a megawatt of power for two seconds at the Joint European Torus (JET) in Oxfordshire on 10 November. By introducing tritium into the plasma, JET scientists demonstrated the capacity of fusion to generate power, but the peak power of 2 megawatts fell well short of the 15 megawatts required to maintain the temperature of the plasma, which is the ‘breakeven goal’ of fusion experiments. Full-power deuterium-tritium experiments are planned for 1996, after which the radioactive experimental apparatus will be de-commissioned.

Roland Pease

Secrecy and the bottom line

Washington & London

EVEN as UK scientists criticized the US National Institutes of Health (NIH) for raising the spectre of commerce in trying to patent gene sequences, the UK genome project has been holding its sequences secret while it prepares to sell them to industry.

Over the past year, researchers working on the Human Genome Mapping Project of the UK Medical Research Council (MRC) have identified about 2,000 human genes, in the form of sequences of fragments of complementary DNA, or cDNA. But rather than publish these sequences or make them available to other scientists, project officials have kept them in closely guarded storage while they developed a database that would allow them to charge industry for access.

Because cDNA, which is the genetic material actually expressed in the body, represents functioning genes, rather than just genetic material that may or may not have a function, researchers consider cDNA sequences more marketable than simple DNA fragments.

The UK project is preparing to sell access to its genome databases at an initial cost of £5,000 per company per year, plus £1,000 for each additional user within the same company. Academic scientists who agreed to sign a non-commercialization agreement would get free access. More than 700 such academics are now using the existing DNA databases of the MRC, which do not yet include the cDNA libraries. No companies have yet been given access to the databases. Although project officials explain that they are simply applying standard MRC policy to "take opportunities for commercialization where they arise", this is believed to be the first time a government has decided to sell its portion of the international genome effort.

"It's well known that the [UK] government does not wish to fund research that is near to the market," says MRC project manager Tony Vickers. "If research has reached a stage of development where an application is round the corner, then industry pays." Vickers believes that cDNA sequences — even though their particular function is still unknown — are close enough to a possible product to justify charging industry for their use.

The UK attempt to sell their genome data "is a sty in the eye of international progress on the genome," says Charles Cantor, chief scientist for the US Department of Energy's Human Genome Project and vice president for the Americas for the Human Genome Organization (HUGO). HUGO plans to convene a meeting soon to discuss how to respond to the MRC

plans. "It's against the spirit of what's being done in the United States," Cantor says. He suspects that criticism from international genome researchers will "force the UK to shift its position on this."

Before settling on the pay-for-access approach, MRC officials had also looked into the possibility of patenting cDNA sequences. But UK lawyers advised them that such patents were unlikely to be granted. Now that NIH has decided to take the patent leap, the MRC group may have been "caught with their knickers down", one NIH researcher says.

If so, this would not be the first time. Behind the MRC preoccupation with commercializing its research is the still-painful memory of the technique MRC scientists developed — and neglected to patent — for producing monoclonal antibodies.

Until Craig Venter of the NIH filed a patent application for 337 cDNA sequences in June (see *Nature* 353, 485; 10 October 1991), the MRC group had intended to open its cDNA library for business this month. Now that NIH has taken the patent route, the MRC group is waiting for a ruling from the US patent office before opening its database in any form. If the patent office finds cDNA patentable, the MRC group expects to seek patents as well. Until that question is answered, however, "all deals are off," says Vickers. "I'd get crucified if I made the sequences publicly available now."

Like many other genome researchers, Vickers has harshly criticized Venter and NIH for attempting to patent the cDNA sequences, something he has described as "seriously prejudicial to the whole thrust of the international Human Genome Project" (see *Nature* 353, 785; 31 October 1991). But NIH scientists are now crying foul: the MRC's plan to keep the data in a closely held database, they argue, runs counter to international agreements to share and exchange genome data.

The difference between the two groups, Venter says, is that his laboratory published (and sought patents on) its data as soon as it was ready; the MRC project has not published its data at all. Venter also says that he has been seeking an exchange of data between the two laboratories for nearly a year, with no success. "We offered on numerous occasions to exchange a complete data set prior to publication and even offered to hold up our submission [to *Science*] so we could publish jointly," he says. "But we were told that the data were to be held secret and would be unavailable for exchange."

Vickers says that the MRC group was not ready to publish when Venter first broached the issue — its sequencing system had been running for just one month.

The MRC group also wanted to get the ground rules for data exchange sorted out internationally before making a start on exchange. Until the United States and other countries had come up with some overall data policy, he says, the MRC group was not prepared to start wholesale transfers of its cDNA sequence library.

But it now appears that the MRC may never publish its cDNA sequences at all. "There is no point in a second publication covering the same ground," Vickers explains. Critics suggest that there may be a commercial reason as well: companies would be unlikely to buy the data if it were publicly available. Whatever the resolution in this case, the end result is clear: now that the issue of commercializing the human genome has become an international controversy, data appear to be the hostage of choice.

**Christopher Anderson
& Peter Aldhous**

DISASTER PREVENTION

Quake raises doubts about Indian dam

New Delhi

AN earthquake that shook northern India on 20 October has added fresh fuel to the controversy over the safety of the \$1,300-billion, 265-metre-high dam under construction in the Himalayas. Tehri Dam, being built with a \$700-million loan and technical help from the Soviet Union, is 100 km from Almora, epicentre of the recent earthquake that killed 1,500 people and shattered 6,000 homes.

The dam suffered no damage from the earthquake, which registered 6.1 on the Richter scale. But environmentalists, led by the Indian National Trust for Art and Cultural Heritage (INTACH) want the work to be stopped to avert a major disaster they say will occur in the event of a bigger earthquake in the fragile mountains. INTACH has filed a petition in the Supreme Court.

Soviet engineers have designed the dam to withstand an earthquake of magnitude 7.2 and a ground acceleration of 0.25 g. But Vinod Gaur, an Indian seismologist, has warned that an earthquake of magnitude 8 or above is likely because of the location in a 'seismic gap' where strain has been accumulating for centuries. He said that such an earthquake may be on the way.

Despite criticisms, the government has shown no signs of giving in to the environmentalist lobby. Prime Minister P.V. Narasimha Rao, returning from a tour of earthquake-ravaged areas, acknowledged that the earthquake has raised some questions about the dam that must be answered. But the ministries responsible for the project say that work on the project will continue without any design changes.

K.S. Jayaraman

Deciding ESA's future

London

MINISTERS from the 13 countries that form the European Space Agency (ESA) will gather in Munich next week to consider Europe's future in space — the first meeting of its type since the Hague conference of 1987 (see *Nature* 330, 195; 1987), when the ESA states (apart from Britain) agreed to support an ambitious manned space programme, including the reusable space vehicle Hermes and Columbus, ESA's contribution to the US-led space station Freedom. But much has changed since 1987, a time of economic plenty across most of Europe, and ESA's director-general, Jean-Marie Luton, must know that his plans to increase spending on its two flagship manned programmes will be opposed by several delegates in Munich.

Luton is seeking a budget that would complete the Columbus programme for about \$5,300 million (at 1990 prices), 14.2 per cent more than the figure agreed at The Hague. This increase is due to delays imposed because of the ESA member states' desire to reduce spending in the early stages of the project (see *Nature* 353, 491; 10 October 1991). The total cost of Hermes, under Luton's proposal, would be about \$7,600 million, some 40 per cent more than agreed four years ago. Again, more than half of this increase is due to the extension of the programme.

Already, signs of discontent are emerging. Norway is expected to announce its withdrawal from Hermes — not a major financial problem for ESA, as Norway contributes only 0.2 per cent of the project's total cost, but just the sort of signal to other wavering nations that ESA's management was hoping to avoid.

The position of the host nation is likely to be a crucial factor in Munich. The costs of German unification have imposed a strict limit on the budget that research minister Heinz Riesenhuber has available for space, leaving him with the difficult job of balancing Germany's future in ESA against national programmes and collaborative projects with the United States and the Soviet Union. The research ministry will not reveal the precise budget ceiling beneath which Riesenhuber has to operate, but the figure is understood to be too low to cover all of Germany's planned space projects.

During the summer, Riesenhuber signalled a tough line when he said that Germany may pull out of Hermes (see *Nature* 352, 177; 11 July 1991), and, says Jan Baldern Mennicken, a senior research ministry official, "that still reflects one of the options." But ESA officials do not expect Riesenhuber to go that far — Hermes is a French-led project, and Germany's withdrawal would be seen as a serious affront

to the French government that could provoke reprisals, damaging ESA projects favoured by the Germans.

Nevertheless, Riesenhuber will be looking for savings in the German contribution to ESA, if Germany's other space interests are not to suffer. And if he fails to find savings in Munich, the meeting's repercussions will spread beyond ESA, threatening the CRAF mission (a collaborative project with the United States to study a comet and the asteroid belt) and a series of collaborations with the Soviet Union, many inherited from former East Germany. Planned collaborations with the Soviet Union now make up almost 30 per cent of Germany's space science spending outside of ESA — about twice the proportion before unification.

Manfred Otterbein, who is responsible for space science at DARA, the German space agency, suggests that any cuts in Germany's space science budget will be a political matter and will reflect the German government's priorities in international relations more than the programmes' scientific content. Mennicken, at the research ministry, disputes this assessment, but political considerations are likely to loom large. Cutting back the proposed German-Soviet collaborations would hit scientists and engineers in former East Germany particularly hard, acting against the German government's current policy of investing in research institutes in the east of the country to bring the eastern scientific infrastructure up to western standards (see *Nature* 352, 267; 25 July 1991). Some savings may be possible, as several Soviet space missions are expected to fall victim to the political and economic uncertainty in that country, but the two main collaborations involving Germany — the Mars 94 and Mars 96 planetary science missions — are thought to be reasonably secure.

This makes the CRAF mission (for which the Germans are supposed to provide some scientific instruments and a propulsion system) look vulnerable. Germany is due to spend some DM120 million (about \$73 million) on CRAF, with more than DM20 million of this planned for 1992–93. But Mennicken says that the research ministry cannot tell if it can afford to contribute to CRAF over the coming year until after the ESA meeting.

A delay in funding CRAF for a year will not be harmful, as the United States is also delaying its part of the programme, but holding off for longer than that could put CRAF in jeopardy. So next week's ESA meeting will be watched closely by US space policy analysts, as well as their counterparts in Europe.

Peter Aldhous

Burt files reopened

London

UNDER pressure from ten of its fellows, the British Psychological Society is reconsidering its 1980 public discrediting of Sir Cyril Burt, the author of controversial publications on the inheritance of intelligence in children.

Burt, who died in 1971, was responsible in the 1930s for pioneering studies demonstrating high correlations between the IQ scores of pairs of monozygotic twins who had been raised in separate homes. But in 1976, the *Sunday Times* published an article alleging that Burt had fabricated data in some of his studies and that two women listed as research assistants for the project had never existed. The British Psychological Society eventually agreed that Burt was guilty of deception in at least three of



Cyril Burt's alleged fraud will be reconsidered.

his publications.

Last month, however, the society's council asked two of its members — Peter Smith, from the University of Lancaster, and Ed Miller, a clinical psychologist at the UK Department of Health — to look again at the Burt affair, to see if the society should formally reopen its investigation. The move followed the request to reconsider the Burt case from the ten society fellows and the publication of two books that sought to clear Burt's name (see *Nature* 340, 340; 1989 & 352, 120; 1991). These argued that any inconsistencies in Burt's published data were the genuine mistakes of an old man, and produced evidence that the missing research assistants did exist.

Bill Wall, a retired child psychologist who has led the campaign within the British Psychological Society to reinstate Burt, argues that his discrediting was politically motivated, coming from psychologists who opposed Burt's ideas on the inheritance of intelligence. On the basis of the latest evidence, "you can disagree with Burt, but you can't say that he was a fraud," Wall maintains. Smith and Miller aim to finish their report by February 1992.

Peter Aldhous

Hopes for Japanese funds fade

Tokyo

THE sudden decision last week by US President George Bush not to visit Japan at the end of this month may kill all hope of a Japanese financial contribution to the Superconducting Super Collider (SSC) project, Japanese government officials say.

Bush was due to meet Japan's new prime minister Kiichi Miyazawa at the end of this month, but he called off the summit and a trip to other parts of Asia, partly in response to criticism in the United States that he has not been giving enough attention to domestic problems.

One of the principal issues to be discussed by Bush and Miyazawa in Tokyo was to have been the SSC. Last month, Japanese government officials told their US counterparts that none of the science-related ministries or agencies of the Japanese government can provide the \$1,000 million or so that the United States is

seeking from Japan to support construction of the SSC (see *Nature* 353, 685; 24 October 1991). The only hope left is that Japan's prime minister may make a political decision to support the project in much the same way as former Prime Minister Toshiki Kaifu agreed to put more than \$9,000 million dollars into the Gulf War.

The November meeting with Bush was ripe for such a decision. Miyazawa, having just taken office, would have been looking for a way to start off on friendly terms with the US president. Also, under Japanese tradition, it would have been important for Miyazawa to give Bush some nice 'souvenir' to take home after their first meeting as leaders of their two countries. The only obvious thing was a contribution to the SSC, say government officials familiar with the planned meeting.

Furthermore, the timing was right. The Japanese budget for next fiscal year will

be finalized at the end of December and Miyazawa could have slipped in an extra provision for the SSC.

The opportunity has been lost with Bush's decision to postpone the meeting. Senior officials from the US Department of Energy will still visit Japan at the end of this month to try to win SSC funds, but an official of Japan's Science and Technology Agency says they do not stand a chance. He says he has repeatedly told US embassy officials in Tokyo that the only hope lies in a Bush-Miyazawa meeting. And, now that that has been postponed, no decision can be made by Japan for at least another year until the next round of Japanese budget decisions, by which time the political climate may have changed completely.

David Swinbanks

UK-JAPAN COLLABORATION

Cold fusion leaves a legacy

Tokyo

THE fuss over cold fusion in 1989 may not have advanced the frontiers of science very much, but it did help to launch a major collaborative effort between UK and Japanese scientists, according to the Japanese leader of the project. The world's most powerful pulsed muon source, on which construction is scheduled to begin early next year, received funding partly because of the early cold fusion claims.

Scientists from the Rutherford Appleton Laboratory in the United Kingdom last week visited the Institute of Physical and Chemical Research (RIKEN) in Wako city near Tokyo to finalize details of plans to build the muon source at the British laboratory with funds from Japan's Science and Technology Agency. The ¥3,000-million (\$23-million) project is one of the first substantial contributions by the Japanese government to British science.

The joint project was first discussed by Japanese and British scientists in late 1988 and it began to move forward a few months later with the support of the British Council in Tokyo. But the key factor that translated an idea into reality was cold fusion, says Kanetada Nagamine of RIKEN, who leads the Japanese side of the project.

In late March 1989, after Stanley Pons and Martin Fleischmann announced their claims of cold fusion, Nagamine was summoned by the science committee of Japan's ruling Liberal Democratic Party to explain what all the fuss was about. His presentation made an impression. Although an earlier request to the Ministry of Education, Science and Culture had been

rejected a few weeks earlier, Nagamine's budget request to the Science and Technology Agency — which had been submitted rather late — began to progress rapidly after the meeting with the committee, and the request to the agency was accepted in July.

Nagamine says this is a "world record" by Japanese standards, because in Japan it normally takes years of "root digging" (*nemawashi*) to launch a project of this size. "It's the only good thing to have come out of cold fusion," he says.

Rutherford Appleton Laboratory was chosen because it has the world's most powerful pulsed proton source, which will be used to make the pulsed muon beam. Construction of the facility will begin in January and the first experiments with the muon beam are expected to begin in 1993 or 1994, according to W.G. Williams, head of the muon group at the British laboratory. The facility, which is funded completely by Japan, will be built by British industry, although the superconducting magnets will come from Japan.

The beam will produce both negatively charged and positively charged high-energy muons. The negatively charged muons, which behave like heavy electrons and are attracted to the positive nuclei of atoms, will be used to investigate muon-catalysed fusion as well as for nondestructive element analysis and for material synthesis through element conversion. The positive muons, which are repelled by nuclei, will be used to characterize materials such as high-temperature superconductors.

David Swinbanks

FRENCH BLOOD SCANDAL

The clot thickens

London

THE controversy surrounding the French blood transfusion service deepened last week with the announcement that another former official has been charged with knowingly supplying infected blood-clotting factors to French haemophiliacs during 1985. The latest focus of the French prosecutors' attention is Jean-Pierre Allain, who in 1985 was head of blood plasma derivatives research at the French national blood transfusion centre (CNTS).

Michel Garretta, former director of CNTS, already faces an identical charge; two other French health officials of the day, Jacques Roux and Robert Netter, are charged with failing to prevent the distribution of the infected clotting factors (see *Nature* 353, 781; 31 October 1991). Central to the charges is the accusation that in 1985 the French health ministry and transfusion service, in order to save money and avoid destroying existing stocks of blood products, allowed the distribution of infected clotting factors instead of swiftly introducing safe, heat-treated supplies.

Allain, who is now a lecturer at the University of Cambridge, protests his innocence: "I spent two years fighting" with Garretta over the need to introduce heat-treated blood products, he says. (In 1985, Allain estimated that 30 to 60 haemophiliacs could become infected with human immunodeficiency virus (HIV) each month, unless the distribution of potentially contaminated clotting factors was stopped.)

Meanwhile, the infighting among former French officials implicated in the scandal becomes more bitter. Roux now faces a libel suit from former French prime minister Laurent Fabius, after alleging that in 1985 Fabius reduced funds for programmes to screen blood for HIV, contributing to the haemophiliacs' plight.

Peter Aldhous

The calm after the storm

London

THE announcement of next year's British science budget, which for the first time will top £1,000 million, has brought a collective sigh of relief from the UK research councils. Following the 1991-92 public expenditure statement, which education and science secretary Kenneth Clarke described as a "real terms level settlement" for research but which the research councils denounced as a cut in science funding, the figure for 1992-93 should slow the rush to cut spending that has occupied the councils for much of the current year. Clarke, speaking at a press conference last week, said the £1,006-million budget represents a real increase of 2.5 per cent over last year's figure, if the government's prediction for the rate of inflation over the coming year holds true.

The heads of the research councils gave the announcement a cautious welcome. "A distinct improvement," said Medical Research Council (MRC) secretary Dai Rees. Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC), which cut back research grants and announced a review of its support for large research facilities after last year's public expenditure announcement, agreed that the budget would allow for a

small amount of growth over this year's spending, but added: "I'm not throwing my hat in the air." The cuts already made by SERC, including the well-publicized decision to close the Nuclear Structure Facility at the council's Daresbury Laboratory in Cheshire, would be unaffected by the new budget, he said.

Tom Blundell, director-general of the Agricultural and Food Research Council (AFRC) welcomed the additional money, but said that this may be too little to cover the £10-million shortfall that AFRC is expecting in 1992-93, mostly because of the difficulty in selling empty laboratory buildings (see *Nature* 353, 690; 24 October 1991). Clarke said last week that it is up to AFRC to convince the Advisory Board for the Research Councils that it needs extra funds. (The board will meet over the coming months to discuss the allocation of the new budget among the research councils). But AFRC "does have a bit of a point," he said.

The MRC's Rees said he was particularly pleased that the planning figures for science spending over the next three years allow for a realistic growth in the budget, taking account of the expected rate of inflation. Traditionally, these planning figures have shown a budget declining in real

terms (a pattern described as "planners' droop" by many civil servants), despite the fact that more money was usually made available later. This has made it difficult for the research councils to plan their future spending. Characteristically, Richmond was more cautious, pointing out that next year will see a general election: the next government may not stick to the planning figures announced last week.

The government's spending plans are seen generally to be a balance between the need to woo voters for the forthcoming election, and the restrictions placed on public spending by the current economic recession. According to the government, spending by the Department and Education and Science has been made a priority, along with the budget of the National Health Service.

Within the education budget, the polytechnics, colleges and universities — soon to become a unified system for higher education (see *Nature* 351, 257; 23 May 1991) — will get 5.5 per cent more in real terms in 1992-93 than in the current year. But given the forecast increase in student numbers, higher-education institutions will be expected to teach each student for a slightly smaller cost. Clarke said he was expecting the higher-education system to produce "efficiency gains" of around 2 per cent.

Peter Aldhous

GLOBAL WARMING

'Activist' stance by UK government

London

THE British government has surprised environmentalist groups with its campaign to encourage homeowners to join its new £10-million, three-year scheme to encourage energy efficiency. Full-page newspaper advertisements published last week juxtapose images of the 'great storm' of 1987 (see *Nature* 329, 750; 1987), which devastated southern England, causing millions of pounds worth of damage and claiming 19 lives, with the stark message: "Global Warming: We have been warned."

"It's leaping ahead of the science," says Simon Roberts, energy campaigner for the British environmentalist group Friends of the Earth, but he adds that he is nonetheless pleased that the government seems to have recognized the need to convince the British public that global warming is an important issue. However, if Friends of the Earth had placed the same advertisements, he observes wryly, "we'd have been torn apart."

To be fair, the text accompanying the images does say that climatologists have not yet proved a link between the excess greenhouse effect and the extreme weather of October 1987, but the public seems likely to overlook the fine print.

The British government's apparent

adoption of an 'activist' stance is in sharp contrast to the position of the US Administration, which has long stressed the scientific uncertainty surrounding predictions of climate change, and it illustrates the deep rift between the countries of the European Communities and the United States over global warming policy. This impasse is hampering attempts to produce a meaningful international climate convention in time for the United Nations 'Earth Summit' in June 1992.

Roberts last week also found himself in the unaccustomed position of welcoming the UK Department of Energy's latest policy on renewable energy sources. Following a Friends of the Earth campaign to increase the emphasis on renewables, the department has announced that it will provide a guaranteed market for 457 MW of electricity generated from renewables between 1992 and 1998 — twice the fig-



**Global Warming.
We have been warned.**

ure it had previously agreed, but still a tiny percentage of Britain's total generating capacity of about 70,000 MW.

Peter Aldhous

Charities fight back

London

BRITAIN's leading medical research charities have decided to take a more active role in the ongoing animal welfare debate. The newly formed Research for Health Charities Group (RHCG), which aims to explain the need for continued use of animals in medical research, is in part a response to two tactics employed by animal rights groups in recent years: a highly successful campaign aimed at schoolchildren, and the attempt to dissuade supporters from donating to charities that fund research using animals.

Myc Riggulsford, who heads the new group, says he intends not only to illustrate the importance of the medical charities' animal-based research, but also to put this work in perspective. The eight charities backing RHCG (which include the 'big three' — the Wellcome Trust, the Imperial Cancer Research Fund and the Cancer Research Campaign) fund a total of more than £200 million of medical research each year — more than the government-funded Medical Research Council — but only £10 million of that is spent on animal research. Riggulsford says that animal rights groups have consistently misrepresented this fact. "We are no longer prepared to let the extremists go unanswered."

Riggulsford dismisses the claim of animal rights groups that medical research could progress solely by using tissue culture and other alternatives to animal experiments. He also points out that most charities that claim to fund no animal research still depend on

animals to some degree. Big C, for example, supports research on cell cultures that grow in a medium containing fetal calf serum — an animal product.

The charities' move is the latest in a series of moves by the British research community to counter the perceived success of animal rights campaigners, who have made a point of targeting young people. Last year, for example, a number of scientific and medical societies, led by the British Association, issued a declaration supporting the continued use of animals in research (see *Nature* 346, 782; 1990). But the new group may have a greater effect, as the medical research charities have a higher profile than most scientific bodies and command greater public support.

Given the activity of groups such as Animal Aid and the British Union for the Abolition of Vivisection in British schools, Riggulsford expects RHCG to focus much of its attention initially on schoolchildren and their teachers, perhaps by offering to provide speakers. The group's supporters would like to see RHCG leading the debate on animal experimentation. "Hitherto, we've just responded to overt or implied criticism from animal rights groups or the media," says Nick Wright, director of clinical research at the Imperial Cancer Research Fund. "We need to be proactive." **Peter Aldous**

ANTARCTICA

New base for South Africa

Cape Town

SOUTH Africa is to embark on major expansion of its Antarctic programme, including the building of a new research station, SANAE 4, and the refurbishment of its polar supply ship, the *SA Agulhas*, to improve its oceanographic research facilities.

The site earmarked for the new base is an inland rock outcrop (nunatak) named Vesleskarvet in the Norwegian sector, about 200 km south of the existing base, SANAE 3, which has reached the end of its lifetime. An environmental impact assessment will be carried out this austral summer, and the base is to be constructed during the 1992–93 season. The new base has been designed to accommodate 80 people during the summer, and a team of 17 to overwinter on the ice.

According to the chairman of the South African Committee for Antarctic Research, Professor Brian Huntley, SANAE 4 will make possible a new research programme on the terrestrial ecology of nunataks. South African research programmes on terrestrial ecology have so far been largely confined to those conducted on Marion and Prince Edward islands, as SANAE 3 has very limited

facilities for biological research.

In addition, the new station will provide vastly superior facilities for the atmospheric physics programme, which is concerned with cosmic rays, stratospheric ozone, and magnetospheric phenomena which are observed at relatively high latitudes.

The refurbishment of the *SA Agulhas* is expected to cost about R20 million (\$7 million) and will begin in April. It will improve facilities for biological, hydrological and geological sampling. These facilities will make possible an enhanced contribution by South Africa to global research programmes on carbon flux and heat flux, in which the Southern Ocean has been identified as a key area.

Apart from the capital expenditure on the new base and refurbishing the *SA Agulhas*, approximately R13 million will be spent annually on running the Antarctic programme for the next five years. This makes Antarctic research very well-funded in South African terms. **Michael Cherry**

BRAZIL

A sad state of affairs

São Paulo

A GROUP of prominent Brazilian scientists is predicting "imminent collapse" of work at universities and research institutions in Brazil because the budget for science and technology "is not being honoured by the federal government". This funding shortfall is not an unusual situation, and each time it happens scientists complain that things are getting worse and worse, so the Brazilian public has learned to turn a deaf ear to these complaints. But this time the scientists may not be crying wolf.

A manifesto describing the situation has been issued by 105 scientists, all of them members of the advisory committee of the National Council for Development of Science and Technology, the main funding agency. The members were elected to the committee by the scientific community, so they are fairly well representative of working scientists.

The manifesto says that 2,316 research projects approved by the committee did not receive any money this year. It adds that even though the year is near its end, only half the budget of another funding agency, the Agency for Financing Studies and Projects, has been delivered. And even if fully funded, that agency's budget amounts to only \$26 million.

Furthermore, the biggest internationally funded programme is paralysed. The World Bank has offered \$150 million to the Programme for the Support of Scientific and Technological Development, but the bank can make the loan available only if the Brazilian government matches it. "This has not yet happened," the manifesto says. **Ricardo Bonalume**

Physicist jumping ship

São Paulo

BRAZIL's pre-eminent physicist, José Leite Lopes, has said he will be leaving the country to return to the University of Strasbourg in France, where he worked in the 1970s during a forced exile while Brazil was in the hands of a military regime. In an angry newspaper interview, he said he is fed up and "disappointed with the whole country" because of such things as rampant corruption, lack of education and children dying of hunger. "I never thought in my youth that we would arrive at this situation, when the country is quickly walking towards the last place in the Third World," he said.

In the late 1940s, full of hope for physics research in Brazil, he founded the Brazilian Centre for Research in Physics with Cesar Lattes, who discovered pion decay in a cosmic ray observatory in the Bolivian Andes. Asked if he intends to return eventually to Brazil, Leite Lopes said he would, at least to visit relatives. "To work, I don't know." **Ricardo Bonalume**

Wrong on animal rights

SIR — Barbara Culliton's article on the animal rights threat to research carries the provocative headline "Can reason defeat unreason?" (*Nature* 351, 517; 1991).

I am constantly amazed at how wrong and even irrational some parts of the scientific establishment can be when it comes to addressing the animal rights argument. Far from adopting rational argument, the research establishment first dismissed the new philosophy as an unrealistic if not crazy message from a group of sentimental misanthropes. By so doing, they grossly underestimated the rational basis of the animal liberation arguments (for example, that developed by Peter Singer or Tom Regan), the strength of public uneasiness about what went on in research laboratories and the Dr Jekyll and Mr Hyde view of science lurking underneath the public's legitimate enjoyment of new health technologies and the latest electronic wizardry.

In the past few years, some major research organizations in the United States have decided that the animal rights community (especially People for the Ethical Treatment of Animals) need to be put back in their proper place (that is, on the fringes of American society). As a result, they have started to tout the benefits of animal research much more aggressively, using high-profile spokespersons such as Michael DeBakey and former Surgeon General Everett Koop. Also, they have portrayed 'animal rights' activists as fringe types who endanger American health standards. In taking these initiatives, their public material has increasingly accentuated appeals to emotion rather than reason in an attempt to counteract the perceived advantage of the animal protection literature. Ironically, the animal activists have been

trying to recast their material to appeal more to reason (including more 'scientific' analyses and more quotations from establishment figures).

While the more aggressive approach of the research establishment will probably have some impact, the arguments and the material miss the mark. First, the US animal protection community has grown from around 2 million to 10 million members in the past ten years. Even though this may represent less than 10 per cent of the adult American population, it shows that there is growing concern about animals. Second, the public accepts the need to use animals in research, so this point does not need emphasizing.

Third, the public is uneasy about the perceived treatment of animals in laboratories and consequent animal suffering. Scientists can most effectively address these concerns by presenting themselves as being equally worried about the moral issues and about addressing the needs of the animals in captivity. To do this, scientists have to accentuate the 'pet-like' nature of laboratory animals (naming them, dealing with exercise and companionship needs and so on) rather than talking of scientifically sound standards of nutrition and the best animal care programme that money can buy. In other words, a scientist must break the standard public stereotype of being a dispassionate, objective and relentless seeker of the truth. Such individuals are respected but are not loved and they are certainly not perceived as good or reliable caregivers.

Recently, a research veterinarian who cares for several hundred chimpanzees, including chimpanzees in AIDS studies, was portrayed on American television as a kindly family physician for the animals.

tists have no choice but to pursue naturalistic explanations for the living world, and scientific explanations can be constructed only from some mixture of random and deterministic processes. Some version of darwinian evolution is currently the only rational scientific explanation for biological life. Darwinian evolution may be an essential component of the atheist's world view, but it is not thereby excluded from the Christian's world view. Pain, evil and the apparent capriciousness of life were acknowledged long before Darwin, and of course the Bible addresses these problems.

CHARLES F. AUSTERBERRY

Department of Biology,
Creighton University,
California Street at 24th,
Omaha, Nebraska 68178, USA

He was shown admiring and tenderly handling a baby macaque. On camera, he said that he would not hire people who said they were not, at least, "bothered" by doing research on animals. On being assured that he was sincere, animal activists said that it would be difficult to criticize studies carried out by such an individual.

The research community is right that their message needs to be imbued with more emotion but they are wrong in thinking that they simply have to show more sick or cured children. Spokespersons have to become much more empathetic as regards the animals they use and show such human characteristics as partiality and emotion in speaking not only of sick humans who need cures but also suffering animals which sometimes help to provide those cures.

Finally, attacking the animal rights message and the messengers is likely to backfire in a country where 80 per cent of the public say that animals have at least some rights — even if this does not include the right not to be killed and eaten.

ANDREW N. ROWAN

Tufts University,
Center for Animals and Public Policy,
200 Westboro Road,
North Grafton,
Massachusetts 01536, USA

Eye-wash

SIR — In a News and Views article (*Nature* 352, 669; 1991), I was quoted as saying at a recent meeting that studying perception by recording from single cells in the central nervous system (CNS) was as misguided as searching for the essence of art in the Louvre.

This is not what I said and not what I believe. Although the Louvre may be a bit over-represented in eighteenth century French painting, it is nevertheless as good a place to seek the essence of art as any I know. Moreover, recording from single cells has taught us a great deal about how the brain functions.

What I actually said was quite different. In an attempt to explain to my more behaviourally orientated colleagues the importance of cellular neuroscience, I said that the study of single cells was the essential foundation for understanding the CNS: for the cognitive scientist, a bit like the men's room of the Louvre, taken for granted but necessary when you need it.

GORDON FAIN

UCLA School of Medicine,
Jules Stein Eye Institute,
Los Angeles,
California 90024, USA

■ The error was made in the *Nature* office. — Editor, *Nature*.

Is God careless?

SIR — As one who accepts both darwinian evolution and orthodox Christianity, I generally appreciated David Hull's review of *Darwin on Trial* (*Nature* 352, 485; 1991). I disagree, however, with his statement that the "God of the Galápagos is careless, wasteful, indifferent, almost diabolical." Must the Creator's characteristics be so simply and directly related to the characteristics of natural processes? The assumption of such a relationship has led to unfounded views of the Universe (for example, as a fair and efficient welfare system) and to unfounded views of God. Neither the Bible as a whole nor the Christ portrayed therein makes such an assumption.

Hull is absolutely correct that scien-

Ornithological cooperation in Siberia

Robert Prýs-Jones

An unprecedented collaborative effort is taking place in a remote region of Siberia to investigate the biology of migratory birds.

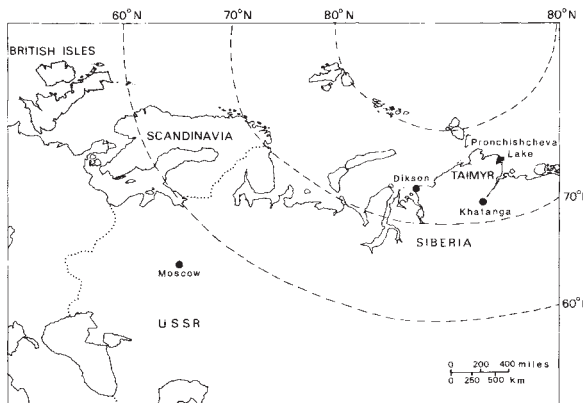
THESE are paradoxical times for co-operative scientific research in the Soviet Union. Politically, possibilities are opening up everywhere, but economic problems are concurrently creating a spreading web of difficulties across the path of sustained studies. Against this background, the International Arctic Expedition, initiated and run since 1989 by Academician E. E. Syroechkovski, head of the laboratory on ecological bases of protection and rational use of fauna at the Institute of Evolutionary Morphology and Animal Ecology (IEMAE), USSR Academy of Sciences, provides a story of unprecedented international interaction in the Taimyr peninsula, a vast and remote area of northern Siberia which was until recently entirely closed to foreign researchers. The focus of this joint work is to improve scientific understanding and to provide more effective conservation of the huge populations of migratory waders and wildfowl that breed there and then fan out along migration routes to wintering areas encompassing western Europe, Africa, southern Asia and Australasia.

A key stimulus to the development of this work has been the desire of Western ornithologists to investigate the causative factors underlying synchronized cycles of breeding success among high Arctic geese and waders, which have been deduced from age ratios observed in wintering populations in western Europe and southern Africa. Although feeding conditions before migration northwards and early summer weather conditions on the breeding grounds may also have some importance, correlational studies drawing on the limited published Soviet information on high Arctic lemming cycles suggest a striking relationship between lemming numbers and the reproductive output of both dark-bellied brent geese (*Branta b. bernicla*) and several wader species, probably mediated by prey-switching and population cycles among arctic foxes (*Alopex lagopus*) and other tundra predators^{1,2}. This research fits well with the broader aims of the Soviet group, which seeks to understand spatial and temporal patterning in the functioning of Arctic tundra ecosystems.

Progress in protection of key migra-

tion and wintering areas in western Europe, notably the Wadden Sea and many British estuaries, has naturally brought about a desire for more integrated protection of key waterfowl sites spanning the entire east Atlantic flyway, including the Siberian breeding areas. Again, these ideas converge with aspirations of Soviet biologists for a Great Arctic Biosphere Reserve, which would greatly extend existing protected areas in the Taimyr to encompass 5 million or more hectares of tundra, continental shelf and islands, as the crown jewel of a proposed chain of reserves within the Soviet Union. The Western Palearctic Waterfowl Agreement should provide a formal international framework for these ambitious goals³.

Studies began in a relatively small way



in 1989, when teams from the then Federal Republic of Germany and Poland, coordinated by WWF-Deutschland and the Gdansk ornithological station, respectively, visited the Taimyr between June and August in conjunction with IEMAE researchers. In 1990, a team from research and conservation institutes in the Dutch Ministry for Agriculture, Nature Management and Fisheries, joined the programme, as well as single researchers from the United Kingdom and from France. By summer 1991, 64 people from 9 countries were participating, including individuals from Finland and Canada and a joint British/South African team comprising three scientists from the British Trust for Ornithology, Royal Society for the Protection of Birds and the University of Cape Town. The participation of a South African researcher is not only encouraging politically but also appropriate biologically, as much data and analysis underlying the

theory linking lemming cycles and arctic wader nest predation originate from South Africa^{4,5}. Although the teams and individuals participating have had diverse research programmes, undertaken at six main research sites widely spread through the Taimyr, there has been considerable coordination in the collection of basic data of mutual interest. At their camp at Pronchishcheva Lake (75° 15'N, 112° 40'E), northeast Taimyr, in an area hitherto almost unknown ornithologically, the British and South African team worked with Dutch and Soviet researchers, and results will be published jointly.

The scale of the logistical problems facing Syroechkovski and his colleagues in carrying forward the International Arctic Expedition, stemming from a combination of the sheer remoteness of the Taimyr and the chaotic economic situation in the Soviet Union, is perhaps difficult for outsiders to grasp. Large quantities of equipment must be taken by air, often involving chartering of planes, from Moscow to the northern Siberian towns of Dikson and Khatanga, before being ferried by helicopter for up to 500 km to field camps. People have to be moved in and out of camps in ways which minimize use of the heavy, long-range helicopters, which are

both indispensable and expensive. Field camps, some in areas previously lacking even basic inventories of fauna and flora, comprise Mongolian-style yurts, circular felted tents heated by diesel-drip stoves, providing warm but basic accommodation. Much ingenuity and planning is used in acquiring both equipment and food for the three-month field seasons, and finance has required approaches to fund-raising, such as sponsorship, which are novel in the Soviet Union. It is to be hoped that the results will demonstrate that the researchers are as efficient as the birds themselves in their international interactions. □

Robert P. Prýs-Jones is in the Estuaries Unit, British Trust for Ornithology, Nunnery Place, Thetford, Norfolk IP24 2PU, UK.

1. Greenwood, J. J. D. *Nature* **328**, 577 (1987).
2. Summers, R. W. & Underhill, L. G. *Bird Study* **34**, 161–171 (1987).
3. Boere, G. C. *ICBP Tech. Publ.* **12**, 345–360 (1991).
4. Underhill, L. G. *Trans R. Soc. S. Afr.* **46**, 209 (1987).
5. Underhill, L. G. et al. *Bird Study* **36**, 83–87 (1989).

Basketball, AIDS and education

The very public announcement of the HIV infection of a famous basketball player in the United States should shame the US government into public education.

New York

TRAGIC though the circumstances are, the announcement last week that 'Magic' Johnson, the outstanding member of the Los Angeles Lakers basketball team, is infected with human immunodeficiency virus (HIV) may yet be the most effective means of getting rid of the confusion that now attends public understanding of this awful and awesome disease. For Johnson has courageously taken the initiative in talking about the predicament in which he finds himself, and has let it be known that his infection stems from heterosexual intercourse. His announcement that he was retiring from basketball because of the HIV infection, delivered in a press conference on 7 November, made headlines across the United States; he subsequently appeared on a popular late-night talk show to discuss his predicament. Johnson's wife of eight weeks, who is newly pregnant, is so far free from HIV, but these are only early days.

Outside the United States, the full drama of these events can hardly be appreciated. Johnson is a hero on the scale of Hercules. In more than a decade on the boards, he has personally made basketball into a fast-moving strategic business, not just an opportunity for competing athletes to show themselves more skillful than their teammates at lobbing a ball into a narrow pipe of netting.

There is more to it than that. Johnson is particularly a hero of the black ghettos, in Los Angeles especially (remember Watts?). Last year, he is reckoned to have earned \$5 million, thus keeping alive the American dream that the poorest kids from the backstreets can also make it if only they are sharp and willing to work. To learn that Johnson has been infected with HIV is a great shock.

How did such a calamity come about? That is what the hero-worshippers ask. Can Johnson be homosexual? He has told them he is not; because he is such a hero, he will be believed without question. And so the truth is now sinking in throughout the United States that AIDS can be transmitted in heterosexual intercourse. Although African experience and the trend of new cases elsewhere has made that clear since 1985 or thereabouts, there remains a stolid subpopulation unwilling to believe. Now, even they are likely to take in the message.

This is an opportunity that the US government should not ignore. Over the past

few years, the first flush of enthusiasm for public education on AIDS has largely evaporated. There is counselling for some of those found to be infected, or to have developed overt AIDS, while the notion of 'safe sex' (which means the use of condoms) is at least widely understood (as is the risk of using drugs intravenously). But voluntary groups are more active than the government, now as always fearful of losing the prurience vote. Johnson's plight seems to have startled the whole United States. Should it not also shame the government into action of a much more energetic kind?

But what can be said that is new, or will carry weight with those at risk of contracting AIDS? Nothing really. But in the United States, where almost everybody knows of somebody who has fallen ill with AIDS, or even died, and where there are many other things to worry people (jobs, for example), it is only natural that the appearance of AIDS in others sometimes seems a proof to many people that it will never happen to them. The novelty is Magic Johnson; if people, especially those most at risk, see that even their heroes can be struck down, it may be easier to bridge that gap in logic. That Johnson wants to help get the message across makes the opportunity more compelling.

The US government has not yet played a significant part in public education on AIDS — a responsibility faced directly by most European and many other governments. Of course, the government encourages voluntary agencies with an inclination to take on this work, while agencies such as the Centers for Disease Control and the National Institutes of Health are even more supportive. But these efforts, however imaginative and well intentioned, cannot compare with a well-run government campaign. The snag is that, as things are, there would be no choice but to plead for a condom culture. There is really no alternative. Vice-President Dan Quayle's statement at the weekend that Johnson's case shows that chastity is the only safe way of avoiding AIDS is correct enough, but suggests that the government is hanging back in the hope that the world will be transformed into a different world, perhaps an ideal one.

What needs to be said (and the need to say it) is entirely unaffected by the continuing uncertainties over the mechanism of the disease. There is no dispute that AIDS is triggered by HIV infection; the

uncertainties arise over what happens then — the pathogenesis of AIDS.

An article in this space a few weeks ago (*Nature* 353, 297; 26 September 1991) showed vividly what passions even that esoteric issue can stir up. The argument, some may recall, was that some data gathered in chimpanzees by a British group and in mice by a Canadian group in British Columbia appeared to be consistent with a view put forward by Geoffrey W. Hoffmann (one of the Canadian group) that the destruction of T cells in AIDS (and pre-AIDS) is caused not by the spread of HIV itself but by an autoimmune reaction between T cells made possible by the similarity of the structure of one of the coat proteins and part of an MHC molecule on the T-cell surface. Unwisely, as it seems in retrospect, the article also mentioned approvingly Professor Peter Duesberg from the University of California at Berkeley, noting that he has consistently argued that more attention should be paid to the curious circumstance that only a small proportion of T cells are found to carry HIV.

The mere mention of Duesberg's name seems to have stimulated both vitriol and misunderstanding. Two other weekly journals, *Science* and *New Scientist*, reported that the article held that Duesberg had been "correct" and that others working in the field disagreed. Duesberg has sometimes argued that HIV is not even a necessary condition for AIDS, but what *Nature's* article said was simply that Duesberg should sleep more easily at night because there is a theory, and some data to support it, to explain the unusual properties of the infected state. Duesberg has made a nuisance of himself, it is true, but does that mean that everything he says and does is wrong?

Otherwise, there have been about a dozen letters (some of which will be published soon) and three offers of more formal papers supporting the Hoffmann view. It will be interesting to see how that comes out.

It might even help Magic Johnson. What would be described, for an ordinary virus, as the incubation period, remains very long. Yet despite efforts everywhere, progress towards a cure is slow, to say the least. Meanwhile, it may be cold comfort for Johnson that his case should make such an important contribution to public education.

John Maddox

Volcanic shade causes cooling

James F. Luhr

PUMICES from the violent 15–16 June eruptions of Mount Pinatubo bore anhydrite (CaSO_4), a mineral found mainly in sedimentary evaporite deposits, Bernard *et al.* report on page 139 of this issue¹. Eruption of the anhydrite-bearing Pinatubo magma injected an enormous SO_2 -rich cloud into the stratosphere, which will probably lead to global-scale cooling at the Earth's surface in the next few years, temporarily counteracting expected trends towards global warming.

In one of the largest explosive eruptions of the century, Mount Pinatubo (Luzon, Philippines) ejected some 3–5 km^3 of dacitic magma² (roughly 10^{16} g) along with a large mass of gaseous sulphur (estimated as 2×10^{13} g of SO_2)³. The remarkably similar 1982 eruptions of anhydrite-bearing magma at El Chichón in Mexico were also accompanied by a large release of SO_2 to the stratosphere. The possibility that volcanic eruptions can modify the Earth's climate has been discussed for more than two centuries, but until about 20 years ago it was erroneously believed that long-lived stratospheric clouds from volcanic eruptions consisted of fine ash or dust. In the early 1970s it was demonstrated that these stratospheric clouds actually consist of micrometre-sized droplets of H_2SO_4 , pointing to the important part played by sulphur in linking volcanoes to climate change.

The behaviour of sulphur in volcanic systems is particularly complex. Depending upon the temperature, pressure and oxidation state of the magma, sulphur can exist as both reduced (sulphide) and oxidized (sulphate) species in four different forms: dissolved in the liquid silicate melt, as an immiscible sulphide melt, in a separate gas phase, and in various sulphide and sulphate minerals. Before 1982, most research into the behaviour of sulphur in volcanic systems focused on reduced conditions, where sulphide species dominate. Experimental studies showed that up to 0.2 wt % sulphur could dissolve in silicate melts under such reducing conditions, appropriate to the voluminous lavas erupted at mid-ocean spreading centres. At the opposite extreme of terrestrial volcanic oxidation states are magmas erupted in subduction zones, where the Earth's oceanic plates, hydrated in contact with sea water, descend beneath overlying plates. The incorporated sea water is driven off when the plate reaches about 100 km depth. This water rises into the overlying mantle and induces melting to form the magmas that

erupt explosively at subduction-zone volcanoes such as Pinatubo and El Chichón.

The resulting high water contents of these magmas are responsible for the explosive nature of the eruptions and the oxidized state of the magmas, which allows crystallization of anhydrite rather than the iron-sulphide mineral pyrrhotite^{4,5}, and causes sulphate (SO_4^{2-}) rather than sulphide (HS^-) to be the dominant sulphur species in the melt⁶. Experimental studies have shown that under such oxidizing conditions, silicate

This coupling may be more than coincidental; several studies have suggested that the El Chichón aerosol could have either triggered or amplified the El Niño^{9,10}. The considerably larger stratospheric aerosol from the Mount Pinatubo eruptions, which is now producing brilliant sunsets in the Northern Hemisphere, has been forecast by Hansen *et al.*⁸ to lead to global-scale cooling of 0.5 °C in 1992–93. This should be sufficiently large to be distinguished from natural variations and from the surface heating associated with the moderate El Niño now underway in the Pacific Ocean. So the Pinatubo eruptions will therefore provide an important test of global climate models.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pinatubo aerosols redden the Hawaiian sunset, 22 August 1991.

melts can dissolve over 0.5 wt % sulphur, more than twice the value for reduced melts^{4,5}. Thus, compared with other classes of terrestrial magmas, hydrous and oxidized magmas formed in subduction zones have a considerably greater ability to store sulphur and to transport it to volcanoes at the Earth's surface, where powerful eruptions can propel it into the stratosphere.

The sulphur gases H_2S and SO_2 , originally present in the magma, are converted in the stratosphere to sulphuric acid aerosols that cool the Earth's surface by backscattering and absorbing solar radiation. Since the emergence of modern global climate models, the El Chichón eruption has provided the only test case for evaluating the climatic influence of a large, volcanically-injected stratospheric aerosol cloud. Cooling of some 0.3 °C at the Earth's surface was forecast to follow the El Chichón eruptions^{7,8}. Evaluation of this forecast was complicated by the relatively small size of the cooling compared to the natural variability in mean global temperature (0.2 °C), but most importantly by the counteracting influence of the great El Niño climate event of 1982–83, one of the most severe this century.

Although the eruptions of Mount Pinatubo ejected considerably larger quantities of both magma (3–5 times more²) and sulphur gases (2½ times more³) than the 1982 eruptions of El Chichón, these two sulphur-rich volcanic events share many intriguing features. Both magmas were very rich in water, as indicated by the abundance of the water-bearing mineral amphibole and by the powerfully explosive nature of the eruptions, which injected material well into the stratosphere and shattered many of the crystals now seen in pumices. Both the Pinatubo and El Chichón magmas were highly oxidized, as reflected by the high values of $\text{Fe}^{3+}/\text{Fe}^{2+}$ in analysed pumices, and by the presence of stable anhydrite crystals. It is worth noting however, that similarly high oxidation states are common for hydrous subduction-zone magmas^{1,5}. So the apparent rarity of anhydrite in subduction-related volcanic sites cannot be attributed to the requirement of an exotically high oxidation state.

More likely, the apparent rarity of volcanic anhydrite is a consequence of two factors. First, the sulphur contents of subduction-zone magmas vary considerably, and most may have insuffi-

Robert S. Fiske

cient sulphur to crystallize a significant amount of anhydrite. Such variability is evident even in comparing the Pinatubo and El Chichón pumices: the latter contain approximately three times as much anhydrite. Equally important, however, is the highly soluble nature of anhydrite in surface waters. Unless completely enclosed by another mineral, anhydrite will not survive more than a few years in pumiceous deposits at the Earth's surface.

Thus, in a cruel trick of nature, the most straightforward criterion for the identification of explosive eruptions of high-sulphur magma is not preserved in the geological record, seriously hampering attempts to identify ancient eruptions of Pinatubo-like, anhydrite-bearing magma that may have modified the Earth's climate in the past. To complicate matters further, an alternative technique of analysing the sulphur contents of glass inclusions trapped within crystals failed badly when used to predict the atmospheric yield of sulphur by the El Chichón eruptions^{5,11}. It is reasonable to expect that the sulphur contents of hydrous subduction-zone magmas might be correlated with some other compositional parameter, but a comparison of the major-element compositions of the Pinatubo and El Chichón pumices shows them to have little in common, save their high sulphur contents¹.

At present, therefore, it is not possible to predict which volcanoes have the potential to erupt high-sulphur, anhydrite-bearing magmas and to produce large, climate-influencing, stratospheric aerosol clouds. Nor can we address this issue for ancient eruptions. Monitoring of sulphur release during eruptive and non-eruptive periods, tracking of the distribution and chemical evolution of stratospheric aerosol clouds, and the compilation of global databases for eruptive activity, gas emission rates, and the composition and mineralogy of eruptive products are all necessary for progress on this front. □

James F. Luhr is with the Global Volcanism Program, National Museum of Natural History, Smithsonian Institution, NHB-119, Washington, DC 20560, USA.

1. Bernard, A., Demaiffe, D., Mattioli, N. & Punongbayan, R. S. *Nature* **354**, 139–140 (1991).
2. Scott, W. E., Hoblitt, R. P., Daligdig, J. A., Besana, G. & Tubianosa, B. S. *Eos* **72**, No. 44, 61–62 (1991).
3. Bluth, G. J. S., Doiron, S. D., Krueger, A. J., Walter, L. S. & Schnetzler, C. C. *Geophys. Res. Lett.* (in the press).
4. Carroll, M. R. & Rutherford, M. J. *J. Petrol.* **28**, 781–801 (1987).
5. Luhr, J. F. *J. Petrol.* **31**, 1071–1114 (1990).
6. Carroll, M. R. & Rutherford, M. J. *Am. Mineral.* **73**, 845–849 (1988).
7. Robock, A. *Geofisica Int.* **23**, 403–414 (1984).
8. Hansen, J., Lacis, A., Ruedy, R. & Sato, M. *Geophys. Res. Lett.* (in the press).
9. Handler, P. J. *Volcanol. geotherm. Res.* **37**, 233–249 (1989).
10. Hirono, M. *J. geophys. Res.* **93**, 5365–5384 (1988).
11. Devine, J. D., Sigurdsson, H. & Davis, A. N. *J. geophys. Res.* **89**, 6309–6325 (1984).

Prediction of progress at last

Janet M. Thornton, Tomas P. Flores, David T. Jones & Mark B. Swindells

THE second fundamental code in molecular biology — how a protein's tertiary structure is coded by its amino-acid sequence — has yet to be cracked. But Benner and Gerloff¹, Bowie *et al.*², and Finkelstein and Reva³ suggest alternative ways forward which offer real hope for progress in the near future. Much of the impetus for these developments has come from the growth in sequence and structure databases.

In the first of the papers, published last year, Benner and Gerloff¹ tackled secondary-structure prediction; this was essentially a case study of the catalytic domain of the protein kinases, the structure of which was then unknown. The cause for excitement is that the structure has since been solved by X-ray crystallography⁴, and Benner and Gerloff's prediction of the core secondary structures was much better than that achieved with standard methods.

Benner and Gerloff were able to use multiple sequence information which, with 79 sequences available for the protein kinase family of structures, proved very powerful. Their approach is not really new^{5,6}, but they illustrate its power when so many sequences are available. This simple method was essentially applied by hand and is as follows. (1) Align all the sequences. (2) Use insertions or deletions, or occasionally special sequences (such as prolines), as parsing points to distinguish loop regions from those of regular secondary structures. (3) Use the pattern of sequence variability at each residue position and the optimal lengths of secondary structures to distinguish between strand and helix.

The results of Benner and Gerloff's predictions are shown in Fig. 1, and are compared to the observed structure and the standard Garnier-Osguthorpe-Robson (GOR) prediction⁷. The parse points are also indicated. Benner and Gerloff correctly identified 13 out of 17 secondary structural elements with over-prediction of only one strand. Of the incorrect assignments, two are terminal secondary structures and there is only one serious mistake in the protein core of the molecule. The power of the method lies in the preliminary parsing of the sequence using insertions and deletions. Of the 16 loop regions, 13 were correctly identified as parse points. However, from their paper we calculated the residue success rate to be only 63%. In comparison, the GOR method applied to the single sequence scores 57%, but only seven of the secondary structures are identified with poor parsing. Thus although the residue-by-residue comparison looks very similar, the possibility of extending prediction to a tertiary fold is much better with the Benner-Gerloff method. The key to their success is the large number of kinase sequences, their remarkable diversity and the accurate hand-derived alignment.

How applicable is this method in general? Insertions and deletions provide remarkably accurate parsing points (a preliminary survey of 45 families indicates better than 90% accuracy) but these are most common in distantly related sequences. Current automated methods cannot accurately align sequences that have less than 35% sequence similarity, without manual intervention

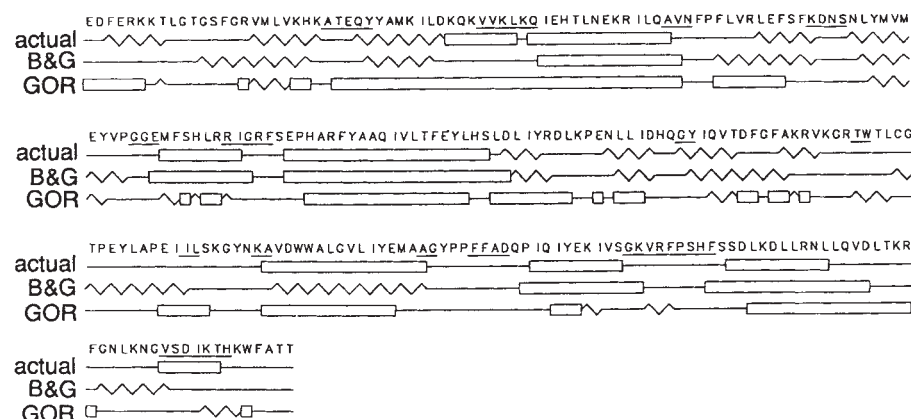


FIG. 1 Observed and predicted secondary structure for the catalytic domain of the protein kinases; single-letter code is used for amino acids. The sequence shown is the β -catalytic subunit of porcine protein kinase for the region predicted by Benner and Gerloff¹ (B & G). Actual structure is taken from the author-derived structural assignments of Knighton *et al.*⁴. Predictions made by Benner and Gerloff are below the actual structure, followed by a single-sequence prediction using the GOR method⁷. Underlined residues indicate regions where insertions and deletions occur in the multiple sequence alignment. Helices are shown as rectangles; strands as zig-zag lines; and coils as straight lines.

and reliable anchor points (known active-site residues for example). In addition, there are few families at present with more than ten sequences available, and their diversity is often limited. At the least, however, Benner and Gerloff's case study provides a hint of what is possible given careful analysis of a large and diverse family of sequences.

The prediction of secondary structure is one path towards the goal of predicting tertiary structure. But as more structures are solved it is apparent that structure, especially topology, is much better conserved than sequence (topology here being taken to be the gross fold and connectivity of the chain as depicted in ribbon diagrams). There are now many examples, in both functionally similar and dissimilar proteins, where the sequence similarity is statistically insignificant (less than 20%) and yet the same topology is observed. So it may be that the stereochemistry of the polypeptide chain imposes a limited number of folds, arising from convergent or divergent evolution. If this is true, the problem of prediction can be rephrased as 'Given a sequence, can we identify the topology?'.

Bowie *et al.*² and Finkelstein and Reva³ have addressed this question, with encouraging results. In essence the problem is one of pattern matching, where instead of matching a sequence with a sequence, a sequence must somehow be matched with a given topology (see Fig. 2).

Bowie *et al.* calculate a table of propensities (three-dimensional/one-dimensional scoring table) which gives the probability for each type of amino acid being found in a given environment. They define 18 different environments, made up of six different accessibility states, each of which may be in the α , β or coil conformation. For a given structure each position can be assigned to one of these 18 states. Dynamic programming is then used to find the best match of the sequence to the pattern of environments found in a given fold. The sequence of the known structure is not used at all in the comparison. This elegant method was tested by taking several specific structures (myoglobin, cyclic AMP receptor protein, ribose-binding protein and actin) and matching them against the complete sequence database. Bowie *et al.* find that those

sequences which are known to belong to the same family or have a similar structure have the highest scores. For example, for actin the structurally similar heat shock proteins were identified, but the more distant yet still topologically similar hexokinases were not.

The results show that this method is better than a database search with a single sequence, because it is less sensitive to specific sequence environments and more sensitive to structural similarity. But the results are only comparable

strands are differentiated) and the preference to be buried or exposed.

Thus, in many respects, this function is similar to the propensities derived by Bowie *et al.* An iterative search of a self-consistent field is used to find the optimal threading of the sequence onto the β -strand lattice and for γ -crystallin the threading produced is close to the native structure. In three other β -sandwich proteins the native folding pattern is among the best three solutions out of 60 possibilities, all other patterns being strictly rejected.

So far, so good. But to go further, the methods will have to be able to recognize distantly related structures such as all the eightfold $\beta\alpha$ TIM barrels, all globin folds including the phycocyanins and the cloverleaf fold of soybean trypsin inhibitor, which is also found in interleukin-1. The problem is that although these structures have the same topology they are sufficiently different that local residue environments have changed.

In a similar study to match folds, Sander and Scharf⁹ have proposed that the accessibility preferences are the crucial factor in distinguishing folds, not the secondary structure preferences (which are rather weak). Taken together with Finkelstein and Reva's results on highly idealized model folds, the implication here is that a detailed consideration of specific interactions may not be necessary. These results emphasize the need for a better theoretical model of the hydrophobic effect. They also show that it may well be fruitful to test a coarse model to simulate protein folding, one based not on detailed residue-residue interactions but just on the burial of hydrophobic groups.

The emergence of new ways of thinking about protein-structure prediction promises a great deal. Even if we are still far from understanding protein folding from energetic principles, we may yet be able to recognize three-dimensional structures from their sequences. □

The authors are in the Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 7HX, UK.

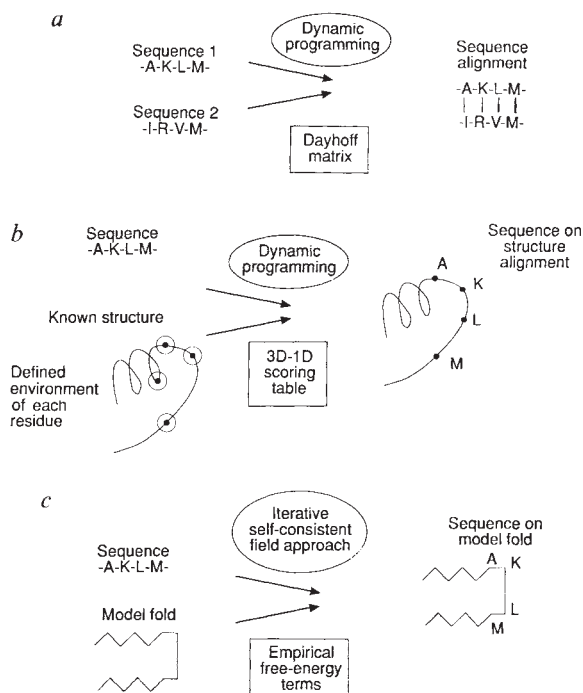


FIG. 2 Approaches to pattern matching. *a*, Sequences with sequences; *b*, sequences onto actual structures (Bowie *et al.*²); *c*, sequences onto model folds (Finkelstein and Reva³). The methods used to score the match are shown in square boxes; the methods to obtain optimal alignment in ellipses.

with searches based on consensus templates derived from the alignment of a family of sequences⁸. In evaluating a knowledge-based method such as this, the procedure should be jack-knifed and the quality of the alignments assessed.

Clearly, the three-dimensional/one-dimensional profile method is limited by the number of different structures in the database. Finkelstein and Reva³ suggest a possible solution to this problem by generating model folds. They describe a method where a sequence is matched against a series of hypothetical folds, in this case, all 60 possible eight-stranded antiparallel β -barrel structures. A lattice representation is used for the model folds which defines exposed and buried residues. The optimal fold is detected using an empirical energy function, which includes terms to represent the preference of an amino acid to be in a β -strand or coil (edge and internal

1. Benner, S. A. & Gerloff, D. *Adv. Enz. Reg.* **31**, 121–181 (1990).
2. Bowie, J. U., Luthy, R. & Eisenberg, D. *Science* **253**, 164–170 (1991).
3. Finkelstein, A. V. & Reva, B. *Nature* **351**, 497–499 (1991).
4. Knighton, D. R. *et al.* *Science* **253**, 407–414 (1991).
5. Crawford, I. P., Niemann, T. & Kirschner, K. *Proteins* **2**, 118–129 (1987).
6. Taylor, W. R. & Green, N. M. *Eur. J. Biochem.* **179**, 241–248 (1989).
7. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).
8. Bashford, D., Chothia, C. & Lesk, A. M. *J. molec. Biol.* **196**, 199–216 (1987).
9. Scharf, M. thesis, Univ. Heidelberg (1989).

Ever decreasing circles

Larry W. Esposito

THE most conspicuous and, until recently, the only known planetary rings in the Solar System, those around Saturn, may be the remains of a comet that came too close to the planet and was torn apart by tidal forces, Luke Dones argues in *Icarus*¹. In doing so, he nearly revives an old suggestion, proposed by Edouard Roche in the last century, that the rings comprise the debris of a moon destroyed in the same way.

The origin of the rings surrounding the planets has, if anything, become a deeper puzzle over the years. In part, this is because, since 1977, rings have been discovered around each of the giant planets Jupiter, Uranus and Neptune, so that Saturn is no longer unique. Moreover, detailed measurements by the Pioneer and Voyager spacecraft have overthrown the old prevailing idea that the rings were formed at the same time as the planets: the evidence shows that gravitational torques and micro-meteoroid erosion quickly destroy the rings, which are typically only 100 million years old, barely 2 per cent of the age — 4.5 billion (10^9) years — of the planets. If the rings are so shortlived, they must be continually replaced for them still be visible.

One possibility^{2,3} is that rings are created whenever a small moon is destroyed in a collision with a comet or asteroid. The moons would themselves be the fragmentary remains of a previous collision involving a larger parent moon. Thus the rings would be the smallest in a cascade of fragments, including unseen tiny moons, orbiting the giant planets⁴. Indeed, Voyager detected a multitude of small moons, including Saturn's recently discovered satellite⁵, Pan, among the planetary rings. And the presence of yet more can be deduced through their effects on the rings⁶. The continuing creation of rings and dust envisaged in this hypothesis would balance the steady loss responsible for the rings' youth.

Although collisional disruption could explain the rings around Jupiter, Uranus and Neptune, Saturn's rings pose a problem. They are so massive that creation by break-up of one of Saturn's moons is very unlikely. The moon would have had to be as large as Saturn's present moon Mimas (with a radius of 193 km): because such large moons are held together strongly by their own gravity, the likelihood is only about 1 per cent that one would have been destroyed in the past 100 million years.

Dones investigates the alternative hypothesis — that Saturn's rings were formed by tidal disruption of a comet

that passed too close to the planet. Within the region now bounded by Saturn's rings, a comet could be broken up by tidal forces and many of the fragments would be captured by Saturn's gravity. The proportion of material captured would depend on the viscosity of the comet, its speed and how close it passed to Saturn. Dones estimates that up to 44 per cent of a comet of radius 500 km could be captured. Within 100,000 years, collisions between the cometary fragments and with Saturn's moons would create smaller fragments

favour of the collisional model. But the tidal model predicts that cometary fragments striking Saturn's inner moons would leave a distinctive crater population, and Dones points out that such a population ('population II') has been observed on the moons Mimas, Dione and Rhea⁷. On balance, we must conclude that the existing observations leave the origin of Saturn's rings undecided.

The tidal model could actually be much more plausible than the collisional model if astronomers have underestimated the number of comets near Saturn. But if comet capture were ten times more likely we would expect that Jupiter, Uranus or Neptune would also have captured a massive ring like Saturn's, as Saturn is not favoured in this

NASA

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A planetary flash in the pan? Saturn's apparently shortlived rings, as captured by Voyager I in 1979. Computer suppression of the three brightest rings allows nearly 100 to be apparent.

whose orbits converge to form a ring.

The model's main drawback is the number of comets approaching Saturn from observations of the object Chiron, whose eccentric orbit crosses those of Saturn and Uranus, Dones predicts that Saturn has had between 10 and 100 close encounters with a comet in the age of the Solar System. Of these, only about one would have left enough debris to form the rings. This formation frequency is similar to that estimated for collisional destruction; in either case, the creation of Saturn's rings is an unlikely event.

In support of collisional creation, Dones notes that the rings are reflective and made of nearly pure water ice; they therefore resemble the inner moons of Saturn and not the dark, dusty comets in their composition. The rings also rotate prograde, that is, in the same direction as Saturn's moons. Because a captured comet has a 50 per cent chance of creating a retrograde (oppositely) rotating ring, this is a weak argument in

particular lottery. Therefore, the present existence of Saturn's rings would still remain unlikely.

Unless we have misunderstood the rapid evolution and short lifetimes of planetary rings (which will be checked by future observation by the Hubble Telescope and the NASA-ESA Cassini mission to Saturn), we must consider ourselves lucky whenever we see Saturn's rings: in the cosmic sense, they are both a rare and a beautiful sight. □

Larry W. Esposito is in the Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309-0392, USA.

1. Dones, L. *Icarus* **92**, 194–203 (1991).
2. Esposito, L. W. *Icarus* **67**, 345–357 (1986).
3. Esposito, L. W. & Colwell, J. E. *Nature* **339**, 605–607 (1989); Erratum, *Nature* **340**, 322 (1989).
4. Colwell, J. E. & Esposito, L. W. *Icarus* **86**, 530–560 (1990); *Geophys. Res. Lett.* **17**, 1741–1744 (1990).
5. Showalter, M. *Nature* **351**, 709–713 (1991).
6. Murray, C. D. & Thompson, R. P. *Nature* **348**, 499 (1990).
7. Lissauer, J. J., Squayres, S. W. & Hartmann, W. K. *J. Geophys. Res.* **93**, 13776–13804 (1988).

Fossil first

THAT hagfish are the most primitive of vertebrates, if indeed they are vertebrates at all, finds compelling support in the first report of a fossil specimen (D. Bardack, *Science* **254**, 701–703; 1991). Apart from a few differences in the position of the branchial skeleton, the 300-million-year-old specimen of *Myxiniroka siroka* from the Upper Carboniferous Mazon Creek fauna in Illinois is similar in most essentials to the modern hagfishes *Myxine* and *Epatretus*. *Myxiniroka* has no trace of any character that might link it with the lampreys or the jawed vertebrates, implying that far from being degenerate forms, modern hagfishes represent a conservative lineage closely related to neither of the other two groups.

Where are quasars?

THE correlation function derived from the positions of quasars on the sky, which measures the departure of their distribution from a purely random one, has the same shape as the galactic correlation function, but is of greater amplitude. Yet the correlation function for rich clusters of galaxies is of greater amplitude yet. This is not coincidence, say N. Bahcall and A. Chokshi (*Astrophys. J.* **380**, L9–L12; 1991). The greater correlation of galaxy clusters follows from the way they sample the underlying population, and the amplitude of the quasar correlation function can then be used to determine the size of the unseen galaxy cluster in which they are presumed to reside. The answer, according to Bahcall and Chokshi, is that optically selected quasars reside in groups of about ten galaxies, and radio quasars in groups of about 30.

Molecular Sisyphus

THE fact that eukaryotic genes can be interrupted by non-coding sequences (introns) allows them to vary greatly in size, independently of their coding potential. So the time taken to transcribe a gene into mRNA also varies greatly, and can be more than the time for a complete cell cycle. Shermoen and O'Farrell (*Cell* **67**, 303–310; 1991) note that transcription stops at mitosis, and have investigated the fate of partially completed transcripts of the long *Drosophila* homeotic gene *Ultrabithorax* (*Ubx*) during early embryonic development when cell cycles are very short. Using a technique that allows them to see nascent transcripts still anchored to their genes in nuclei, Shermoen and O'Farrell find that the transcripts are aborted at mitosis. Thus, Sisyphus like, transcription of genes as long as *Ubx* must start from scratch at the beginning of each of the short cell cycles, and full length mRNA cannot be made until the cell cycles lengthen.

Temporal associations

Michael P. Stryker

THE temporal lobe of the brain has long been considered to have a crucial role in both memory and visual imagery¹; the selective responses of inferotemporal cortical neurons to complex objects, such as hands or pictures, or models of monkeys' faces, have intrigued neuroscientists for nearly two decades^{2–4}. A report from Sakai and Miyashita on page 152 of this issue⁵ now reveals that neurons in the temporal neocortex of macaque monkeys can be taught to associate pairs of arbitrary geometrical stimuli. This capacity of temporal cortex for making associations may be the key to understanding its other phenomenal capacity, the ability of temporal cortical neurons to respond similarly to the different views of an object that result from movement of the object or the observer.

Part of the fascination of this area of the brain is that it has been considered by many to contain the highest stage of refinement of visual receptive fields. In this view, visual features that trigger neurons were thought to become more and more specific as one progresses from the primary visual cortex V1 through successive stages of visual information processing. In V1, neurons respond well to simple light–dark edges of particular orientations at particular places within the visual field. But by the stage of the temporal cortex, it was suggested, neurons would be so selective that individual ones might respond only to the image of one's grandmother, and the discharge of such 'cardinal' neurons would be taken by the rest of the brain as evidence that one's grandmother was in view⁶. Temporal lobe recordings that demonstrated highly selective responses to complex visual stimuli^{2–4} were consistent with this view.

One of the difficulties for this account of temporal-lobe neuronal responses is that in many cells such responses persist independently of the particular point of view of the optimal stimulus. How could a simple visual pattern-recognition machine respond in the same way to grandmother's full frontal image as to her profile? One solution offered is to posit a hierarchy of detector neurons specific for different components of the image of, for example, a face. The face-selective neuron would then respond only if it received simultaneous input from some eyebrow neurons, nose neurons and mouth neurons⁴. As the appearance of the face changes so dramatically with point of view, it was unclear whether this account was satisfactory, or whether a number of cardinal views of a face would have to be stored

as separate conjunctions of trigger features. Another proposed solution is that the visual system may either reconstruct object-centred three-dimensional representations from two-dimensional views^{7–9} or else incorporate computational machinery for determining whether different views are of the same scene, a problem for which attractive algorithms have been described¹⁰.

Three years ago, Miyashita's laboratory¹¹ reported a finding from the temporal cortex similar to those made in the hippocampus and in other higher cortical areas¹²: when the responses of temporal cortical neurons to a large set of complex visual stimuli were measured, individual neurons commonly responded well to only a small fraction of the stimuli that were presented, and different neurons responded to different selections from among the stimuli. Miyashita's group¹¹ looked for similarities among the visual stimuli that elicited good responses in temporal cortical neurons. In order to avoid complications due to familiarity or significance of the stimuli, they used novel computer-generated fractal images which the animal subjects could not have seen previously. They found no geometrical similarities between the effective stimuli; instead, the only similarity among the particular collections of stimuli selected by different neurons was that they were presented sequentially (or nearly so) during training trials, in which the same series of 100 fractal stimuli were presented in the same order over and over again¹³. This suggested that the temporal cortex had learned the serial order of the stimuli, and had somehow associated (in the sense that the neurons responded similarly to) consecutively presented stimuli with one another. The work of Sakai and Miyashita⁵ is a direct test of that hypothesis. Its results confirm the

1. Iwai, E. and Mishkin, M. *Vision, Memory, and the Temporal Lobe* (Elsevier, New York, 1990).
2. Gross, C. G. et al. *Science* **166**, 1303–1306 (1969).
3. Desimone, R. et al. *J. Neurosci.* **4**, 2051–2062 (1984).
4. Perrett, D. I. et al. *Trends Neurosci.* **10**, 358–364 (1987).
5. Sakai, K. & Miyashita, Y. *Nature* **354**, 152–155 (1991).
6. Barlow, H. B. *Perception* **1**, 371–394 (1972).
7. Perrett, D. I. et al. *J. exp. Biol.* **146**, 87–112 (1989).
8. Marr, D. *Vision* (Freeman, San Francisco, 1982).
9. Hasselmo, M. E. et al. *Expl Brain Res.* **75**, 417–429 (1989).
10. Ullman, S. *Cold Spring Harbor Symp. quant. Biol.* **55**, 889–898 (1990).
11. Miyashita, Y. & Chang, H. S. *Nature* **331**, 68–70 (1988).
12. Rolls, E. T. in *The Neural and Molecular Bases of Learning* (eds Changeux, J.-P. & Konishi, M.) 503–540 (Wiley, Chichester, 1987).
13. Miyashita, Y. *Nature* **335**, 817–820 (1988).
14. Poggio, T. *Cold Spring Harbor Symp. quant. Biol.* **55**, 899–910 (1990).

idea that temporal-lobe neurons can learn to associate stimuli on the basis of temporal contiguity. It raises the possibility that such an associational mechanism may be a general property of this area of cortex.

The empiricist philosopher David Hume said that causality in nature is nothing more than constant coincidence. In our experience of the visual world, the different views of an object are nearly always presented in succession, transformed by the movement of the object or the observer. If the temporal cortex contains a general mechanism for the association of successively presented stimuli, then these views would automatically become associated, obviating the need for special mechanisms for doing geometrical transformations or for hierarchically associating conjunctions of trigger features. The computational

geometry performed by the temporal lobe may reflect nothing more than the operation of a neural system set up to detect coincidence in nature.

This hypothesis, like the present experiment, is also attractive because it unifies the perceptual and limbic aspects of temporal neocortex by positing that the visual properties of temporal cortex may be a consequence of its essential limbic, memory-related capacity. It may be an example of how the brain makes the trade-off between memory and computation¹⁴ in the solution of a problem of perceptual constancy. □

Michael P. Stryker is in the W. M. Keck Foundation Center for Integrative Neuroscience and the Neuroscience Graduate Program, Department of Physiology, University of California, San Francisco, California 94143-0444, USA.

CELL DIVISION

Ring of bright metal

Ken Begg

THE most benevolent commentator could hardly claim that advances in the field of bacterial cell division have been rapid; the more cynical might suggest that reviews on the subject have increased in inverse proportion to progress. But the question remains important, and study of division in *Escherichia coli* still represents our best chance of

daughter cells with almost unerring accuracy.

Early attempts to investigate these events centred on whether cell-envelope growth progressed by incorporation of new material at specific growth zones or by random intercalation of components all over the surface. The idea of zonal growth was attractive because it offered a differentiated structure in which to place a septum. It also provided a mechanism for chromosome segregation if, as was suggested in a classic model², they were attached to membrane. Despite much ingenious experimenting, however, all we have is a wealth of contradictory evidence and no clear answer.

Genetic studies on temperature-sensitive mutants that fail to septate at the restrictive temperature have identified several genes associated with cell division and a degree of sequential order has been assigned to

them. There is disagreement about the order in which they operate, but general consent that *ftsZ* is one of the most important of them and that it is required at the earliest known step of septation^{3,4}. The gene is the target for the 'natural' division inhibitor of the MinCD system, elegantly explained by de Boer *et al.*⁵, and is also the target for several other inhibitors including the stress-induced protein, SulA. In addition, the level of *ftsZ* protein (FtsZ) regulates the fre-

quency of division⁶ and the protein is widespread throughout the eubacteria⁷.

Much of our understanding of this gene and its product comes from Lutkenhaus's laboratory in Kansas City. He and Bi now report that FtsZ is cytoplasmic; it is dispersed in the youngest cells but aggregates to form a ring-like structure when a critical level is reached. They propose that the protein self-assembles into a dynamic ring on the inner surface of the cytoplasmic membrane at the place where division will occur, and that the formation of the ring is the signal for septation to begin (see their Fig. 3 on page 163 of this issue). Their results, based on immunoelectron-microscopy using immunogold, are clear and convincing.

Bi and Lutkenhaus make no attempt to correlate their findings with the other reported ring-like structure, the periseptal annulus, but the matter deserves some thought. MacAlister *et al.*⁸ reported the presence of this division-related organelle, and studies in Rothfield's laboratory in Farmington, Connecticut, have led to the proposition that these annuli denote future division sites. Annuli are found at the mid-point of even the youngest cells; they replicate, are displaced laterally and are then arrested in new locations by some unknown signal. In a recent review⁹, de Boer *et al.* mention, as unpublished observations, that nonseptate filaments of the temperature-sensitive *ftsZ84* mutant fail to localize periseptal annuli correctly. They assign this mutant to a class which might "fail to arrest the displacement process when the annuli reach their proper locations". This would fit well with what Bi and Lutkenhaus propose. Assembly of the FtsZ ring might be the signal that stops lateral displacement of the periseptal annuli, therefore initiating the process of septation.

It is interesting to note that the criteria Bi and Lutkenhaus impose in screening for septal section, and their finding that sections containing DNA show randomly dispersed FtsZ, suggest that nuclear segregation has already taken place by the time the ring starts to form. The authors mention the possibility of a link. But although no one would disagree that

One goes into two — a false-colour image of *Escherichia coli* in the process of undergoing cell division.

success. Indeed, promise is beginning to turn into reality, as Bi and Lutkenhaus show on page 161 of this issue¹. They present the first report of a protein specifically localized to the division site.

Escherichia coli is a rod-shaped organism which doubles in length over the course of a generation time, and then divides regularly and precisely across its middle. During the course of this cycle it will replicate its DNA and partition the resulting chromosomes into the two

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

1. Bi, E. & Lutkenhaus, J. *Nature* **354**, 161–164 (1991).
2. Jacob, F., Brenner, S. & Cuzin, F. *Cold Spring Harb. Symp. quant. Biol.* **28**, 329–348 (1963).
3. Begg, K. J. & Donachie, W. D. *J. Bact.* **163**, 615–622 (1985).
4. Taschner, P. E. M. *et al. J. Bact.* **170**, 1533–1540 (1988).
5. de Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. *Cell* **56**, 641–649 (1989).
6. Dai, K. & Lutkenhaus, J. *J. Bact.* **172**, 3500–3506 (1991).
7. Corton, J. C., Ward, J. E., Jr & Lutkenhaus, J. *J. Bact.* **169**, 1–7 (1987).
8. MacAlister, T. J., MacDonald, B. & Rothfield, L. I. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1372–1376 (1983).
9. de Boer, P. A. J., Cook, W. R. & Rothfield, L. I. *A. Rev. Genet.* **24**, 249–274 (1990).

chromosome segregation and septation are well coordinated in the normal cell, there is too much evidence against chromosome segregation being a necessary prerequisite for division.

Bi and Lutkenhaus have demonstrated the value of electronmicroscopy using immunogold in assigning a role to a key gene product. With this approach, and given the renewed interest in nuclear

segregation in both Europe and Japan, we can expect to see more progress in the next five years than we have in the past twenty. □

Ken Begg is on secondment from the Medical Research Council in the Institute of Cell and Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK.

OPHIOLITES

Ocean floor comes ashore

Julian Pearce

ON page 140 of this issue¹, Don Elthon discusses the composition of the Bay of Islands ophiolite complex in western Newfoundland. His observations, and related work in press by Jenner and coworkers², highlight the problem of finding exact on-land analogues of the crust and mantle created at mid-ocean ridges.

Ophiolite complexes are fragments of oceanic crust and upper mantle detached and uplifted from their original deep-ocean setting by global tectonic processes. Land-based geologists recognized their significance at about the same time that dredging and geophysical studies of the mid-ocean ridge system led to the development of the sea-floor spreading and plate tectonic hypotheses³. Their results prompted a Penrose conference in 1972 to produce a formal definition of an 'ophiolite'. According to the conference participants⁴, a fully-preserved ophiolite complex is a layered assemblage of mafic to ultramafic rocks which has mantle peridotite at its base, then a crustal sequence composed mainly of plutonic rocks overlain by sheeted dykes, pillow lavas and pelagic sedi-

ments. The sheeted dykes, continuous exposures of dykes cutting dykes, indicate infinite extension and so provide the principal evidence for an origin by sea-floor spreading⁵. Thus, these complexes potentially provide a great opportunity to study the deep structure and composition of the oceanic lithosphere without getting wet and without spending the millions of dollars entailed in deep-sea drilling projects such as Mohole.

The first doubts about whether ophiolite complexes are typical of the lithosphere formed at mid-ocean ridges were raised, amid much controversy, by Miyashiro⁶, who pointed out that rocks from the 'type' ophiolite, the Troodos Massif of Cyprus, differed significantly from rocks from the ocean ridges, having for example a higher average content of silica and a less iron-rich crystallization path. Such compositions, he maintained, were more typical of the rocks erupted from volcanoes above the subduction zone in his native Japan.

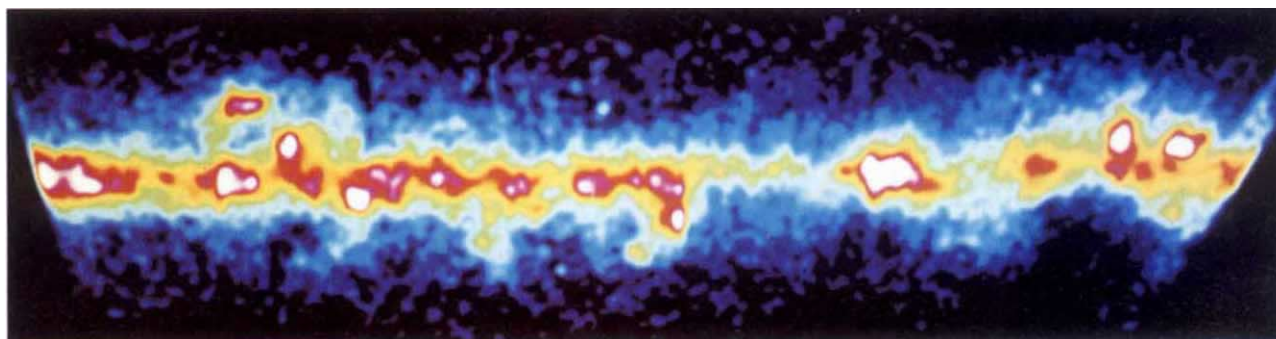
Since then, more and more ophiolites have been placed under the geochemical microscope and found to have the composition of subduction-zone, rather than

mid-ocean-ridge, volcanic rocks. Geologists have since reached a broad consensus that a significant proportion of the world's ophiolite complexes formed above subduction zones, though not in island-arc volcanoes as Miyashiro had envisaged, but at ocean ridges in the small basins that form within and around these volcanic chains⁷. Nonetheless, a small number of well-preserved ophiolites retained their status as having formed at true mid-ocean ridges. Principal amongst them was the Bay of Islands ophiolite, on whose progeny Elthon now casts doubt.

Elthon uses the trace-element abundances in the lavas, dykes and uppermost plutonic rocks as geochemical fingerprints of the tectonic setting of formation of the Bay of Islands ophiolite. Because the rocks have undergone metamorphism, he restricts his choice of elements to those of high ionic potential (charge/radius ratio) that are most likely to have remained immobile during alteration. Two key elements are thorium and tantalum. These elements are in similar concentration (about 0.02 parts per million) in the mantle asthenosphere that rises and melts beneath mid-ocean ridges. They are also partitioned to a similar degree between melt and the main rock-forming minerals so that the resulting ratio in mid-ocean ridge basalts remains close to unity.

In subduction zones, however, the larger thorium ion is more readily taken up by the aqueous and siliceous fluids that are driven off the subducted oceanic crust and its sedimentary cover into the overlying wedge-shaped volume of mantle. Melting of this enriched mantle wedge then gives magma with anomalously high Th/Ta ratios. Elthon¹ shows that the Th/Ta ratios in the Bay of Islands ophiolite range from about 3 to

Going to great wavelengths in the Milky Way



THIS new false-colour image of the Milky Way shows not the stars in our Galaxy, but the gas where new stars are destined to be born. The image was obtained by a purpose-built, balloon-borne telescope, constructed by a Japanese-US team, which was tuned to detect solely the infrared radiation from ionized carbon (seen at the edges of dense, star-forming

clouds, and in the picture showing as intense red-yellow knots). Ionized carbon may play a crucial part in both helping and hindering star formation, by radiating heat away from the collapsing clouds of gas, and by injecting turbulence into them through its interaction with magnetic fields.

ROLAND PEASE

H. Okuda, Institute of Space and Astronautical Science

5, well above the values of 0.75–2 measured in lavas from the present-day mid-ocean ridge system. The La/Ta ratio, a similar indicator, has values of 30–40 in the Bay of Islands compared with 10–20 in mid-ocean ridges. When tantalum is plotted between thorium and lanthanum on a multi-element pattern it gives the negative anomaly that is commonly regarded as diagnostic of a subduction-related origin. Jenner *et al.*² use niobium in place of tantalum, coming to similar conclusions.

Thus, Elthon's work again raises the question of whether any well-preserved ophiolites can be ascribed to a mid-ocean-ridge origin. The problem appears to be one of preservation. Although the volume of lithosphere created at mid-ocean ridges exceeds that formed at small spreading centres above subduction zones by several orders of magnitude, it is dense and tends to be subducted into the mantle. By contrast, oceanic lithosphere created above a subduction zone may be buoyed up because the underlying mantle has been hydrated (serpentinized) by water driven off the subduction zone. It may also be directly underthrust by the continental margins that eventually reach the subduction zone, leading to its preferential detachment and uplift. Ophiolites with geochemical signatures indicative of a mid-ocean-ridge origin tend to be preserved as dismembered fragments of lithosphere that were scraped off a subducting plate or entrained in a continent collision (many complexes in the Alps are examples of this). The few that are well-preserved (for example, Macquarie Island south of New Zealand) occupy anomalous tectonic settings.

Understanding ridge-crest processes is important not just for the intellectual satisfaction of understanding the creation of a significant part of the Earth's surface, but because the hydrothermal circulation that gives rise to the high-temperature metal-rich exhalations known as 'black smokers' has economic potential and because the interaction of sea water with oceanic crust contributes to the global budget of carbon and hence of greenhouse gases. Of course, the three-dimensional exposure provided by ophiolites means that the study of these on-land outcrops will remain essential to our understanding of ridge processes, but the fact that lithosphere created above subduction zones has a distinctive

chemical and mineralogical composition implies that ophiolites and 'normal' oceanic crust both have to be studied.

This absence of good ophiolitic analogues for 'normal' oceanic lithosphere has resurrected the concept of the Mohole drilling project, albeit in modified form. A primary objective of the multi-national Ocean Drilling Program is to drill the deeper parts of the actual

oceanic crust and upper mantle in places where they are exposed near the sea floor. That the rocks recovered will have the same composition as those from most ophiolite complexes now appears even more unlikely. □

Julian Pearce is in the Department of Geological Sciences, University of Durham, Durham DH1 3LE, UK.

ANTHROPOLOGY

Why are pygmies small?

Jared M. Diamond

OVER 4,000 years ago, the Egyptian Pharaoh Nefkare instructed his soldiers to take good care of a "dancing dwarf" captured near the Upper Nile and being brought back alive to the Pharaoh's court. That letter is the oldest preserved reference to African pygmies. But at what age, through what hormonal mechanisms and because of what evolutionary factors do pygmies become smaller than other humans? A longitudinal study of growth in pygmies of known birth date, carried out by Robert Bailey and just published in *Annals of Human Biology*¹, allows speculation on those questions to be grounded more firmly in fact.

As background, consider the problems in measuring growth rates of nomadic, illiterate people who lack a calendar and whose age cannot be safely inferred from comparisons with European norms. The most detailed previous attempt to solve these problems was a cross-sectional study by van de Koppel and Hewlett², who measured African pygmies whose ages were estimated by asking parents to place their child's birth in time relation to dated events (for instance the death of an important chief). They concluded that pygmies are of normal size until puberty, whereupon they fail to undergo the adolescent growth spurt characteristic of other humans. As a possible physiological mechanism, Merimee and Rimoin³ reported that pygmies are normal in their secretion of growth hormone but abnormally unresponsive to that hormone and that they have abnormally low levels of insulin-like growth factor. These results would be important not only in providing developmental and physiological explanations of pygmy size, but more broadly in implying that insulin-like growth factor is what triggers the normal human growth spurt at puberty.

In 1980 Bailey began his long-term study, in Zaire's Ituri forest, of the Efe pygmies, the world's smallest people (mean adult height 1.42 m (4 ft 8 in) for men, 1.35 m (4 ft 5 in) for women). He and his colleagues recorded birth date,

weight and height for 51 Efe children and 52 children of their black African farmer neighbours, the Lese. They then remeasured the height of as many subjects as possible every six months over the first five years of life.

It turns out that even at birth, with a mean weight of only 2.7 kg, Efe infants are smaller than those of the Lese. Efe height falls progressively further behind up to at least age five. In contrast, although even the Lese are small by US standards (third percentile for height), the Lese maintain that proportionate size deficit approximately constant through to age five. No Efe child, though, surpasses the US third percentile after age three months, and the average height of a five-year-old Efe is the same as that of a 2½-year-old US girl. These differences between Efe and Lese surely involve a large genetic component, because the Efe seem to be no worse off than the Lese with respect to nutrition and parasites.

Even if the uncertain age estimates of pygmies in early cross-sectional studies were correct, van de Koppel and Hewlett's graph of growth (page 97 of ref. 2) appears on reinspection to be compatible with some growth spurt at puberty. In Bailey's study, most or all of the Efe pygmies' proportional deficit in height had been reached by age five; so they would still end up small even if they did experience a normal adolescent growth spurt. Should this developmental explanation of pygmy size prove correct, it will warrant reconsideration of hormonal theories of the growth spurt.

What ultimate selective factors programmed the genetic basis for Efe body size? There is a large zoological literature about intra- and interspecific differences in animal body size, and there prove to be numerous factors involved, including effects of size on heat control, competitive rank and energy requirements. Among these factors, Cavalli-Sforza⁴ has pointed out why thermal considerations would be expected to favour small body size in the hot, humid climate of a tropical forest. In that

1. Elthon, D. *Nature* **354**, 140–143 (1991).
2. Jenner, G. A., Dunning, G. R., Malpas, J., Brown, M. & Brace, T. *Can. J. Earth Sci.* (in the press).
3. Gass, I. G. *Nature* **220**, 39–42 (1968).
4. *Geotimes* **17**, 24–25 (1972).
5. Kidd, R. G. W. & Cann, J. R. *Earth planet. Sci. Lett.* **24**, 151–155 (1974).
6. Miyashiro, A. *Earth planet. Sci. Lett.* **25**, 217–222 (1975).
7. Pearce, J. A., Lippards, S. J. & Roberts, S. *Geol. Soc. Lond. Spec. Publ.* **16**, 17–94 (1984).

environment there is a big risk from exhaustion due to the body's own heat production during exercise (high humidity makes heat loss by sweating inefficient). That risk would be minimized by the small body size (hence low mass: area ratio) and relatively small muscle mass of African pygmies.

That reasoning must, however, be placed in the context of human size

getting through dense vegetation. The lowland rainforest of New Guinea is open enough for even basketball players to walk erect, but the montane forest is a tightly woven maze, as I am reminded every time I bang my head trying to crawl after a little Papuan mountaineer. Turnbull⁷ noted this advantage for forest-dwelling African pygmies as well, and Bailey observed that Efe men spend

nine per cent of their waking hours clambering around in trees in search of bees' nests with honey — not a safe lifestyle for a sumo wrestler.

Yet this advantage of small size in dense vegetation fails in turn to account for the Kalahari Bushmen, who live in an open habitat. The explanatory factor there may be the desert's low and fluctuating productivity: small people are better able to endure starvation under such conditions. (Recall that the first man to die during Robert Scott's fatal dash to and from the South Pole was big Edgar Evans, starved by months of democratically equal division of dwindling food supplies⁸.) The low productivity (in wild game and agricultural protein) of New

Guinea's mountains may similarly favour small statures, and may also contribute to the size of African pygmies; Bailey points out that although they are often portrayed as hunters at home in the forest, African pygmies recognize that they would starve there and actually obtain most of their food by working for nearby farmers^{5,6}.

Pygmy stature has intrigued scientists for its own sake, just as it intrigued the Pharaoh Nefrikare, Homer and Herodotus. Bailey's work points towards further studies to go beyond that and illuminate the whole issue of what determines human body size. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

1. Bailey, R. C. *Ann. hum. Biol.* **18**, 113–120 (1991).
2. van de Koppel, J. M. H. & Hewlett, B. S. in *African Pygmies* (ed. Cavalli-Sforza, L. L.) 95–101 (Academic, New York, 1986).
3. Merimee, T. J. & Rimoin, D. L. in *African Pygmies* (ed. Cavalli-Sforza, L. L.) 167–177 (Academic, New York, 1986).
4. Cavalli-Sforza, L. L. in *African Pygmies* (ed. Cavalli-Sforza, L. L.) 361–426 (Academic, New York, 1986).
5. Bailey, R. C. & Peacock, N. R. in *Coping with Uncertainty in Food Supply* (eds de Garine, I. & Harrison, G. A.) 88–117 (Clarendon, Oxford, 1988).
6. Bailey, R. C. et al. *Am. Anthropologist* **91**, 59–82 (1989).
7. Turnbull, C. M. in *African Pygmies* (ed. Cavalli-Sforza, L. L.) 103–123 (Academic, New York, 1986).
8. Scott, R. F. *Scott's Last Expedition* (Smith, Elder, London, 1913).



Small world — the Efe pygmies are the world's smallest people, the mean adult height for men and women being 4ft 8in and 4ft 5in respectively. (Photo courtesy Robert C. Bailey.)

variation worldwide. Although African pygmies are the smallest of humans, they have many close rivals in certain populations in tropical America, southern Africa and (especially) Asia, the Philippines and New Guinea. Of these populations, the only ones living in hot, humid forests are the Andaman islanders and some Amerindians. In striking contrast to predictions based on Bergmann's rule and heat budgets, human body size consistently declines (rather than increases) with elevation in New Guinea, where the cold, high mountains harbour the smallest people (such as the Goliath Pygmies, oxymoronically named after their homeland on Mount Goliath). The small negritos of the Philippines and the Semang of Malaysia also live in the mountains, not in the steaming lowlands. Kalahari Bushmen are small despite living in a seasonally hot but extremely dry climate. Even the pygmies of equatorial Africa may have evolved at the productive savanna/forest ecotone and been driven into the humid forest only by the Bantu expansion of the past 3,000 years^{5,6}. These many paradoxes suggest that there must be evolutionary explanations of pygmy size that are not based solely on thermal considerations.

One such advantage of small size in cool equatorial mountainous areas is in

Bacterial choirs

MANY biochemical changes go by means of a closed cycle of reactions. Each cycle takes in reagents, liberates products, and regenerates the intermediates to start itself again. Thus glucose is oxidized in many organisms via the famous Krebs citric-acid cycle. Daedalus now points out that all chemical reactions go with a small change of volume; for a cyclic reaction this change must be cyclic. The tissue supporting a biochemical cycle must expand and contract once per cycle. So it should generate sound: a hum or whistle or ultrasonic signal, depending on how fast the cycle is running.

Living tissue does not obviously whistle while it works. Daedalus explains that any given biochemical cycle operates simultaneously at many sites in each cell or mitochondrion. All the sites have different frequencies and phases, so that their total sound output averages to zero. Daedalus is now synchronizing them.

He argues that a chemical reaction that produces an increase in volume must be slowed by a rise in pressure, but accelerated by a fall. Conversely, a reaction going with a decrease in volume will be encouraged by added pressure and slowed by its reduction. A periodic fluctuation of pressure imposed on a cyclic reaction will therefore tend to entrain that reaction. Its expansion phase will be trapped in the periods of low pressure, while its contraction phase fills the periods of high pressure.

So DREADCO biochemists are now singing to bacteria. They are imposing specific sonic frequencies on bacterial cultures, calculated to match, for example, the Krebs cycle. When the correct frequency and intensity have been found, all the cycles in the culture will come into coherent phase. The bacteria will sing back in close harmony.

Even more intriguing, they should continue to sing when the entraining frequency is switched off. Like a laser, they will maintain their own synchronizing signal, and will function as a coherent free-running oscillator. Thereafter, their frequency will be a perfect monitor of their metabolic rate.

This elegant technique has wide implications. Almost any living tissue, in plants, animals or human beings, could have one of its biochemical cycles entrained and synchronized in this way, and could thereafter sing its reaction rate into a suitable microphone. At last biochemists will have instant, second-to-second knowledge of the changing metabolic rate of individual tissues. They will sing of their own health, sickness and response to treatment — even, by a note sinking sadly to zero frequency, of their own death.

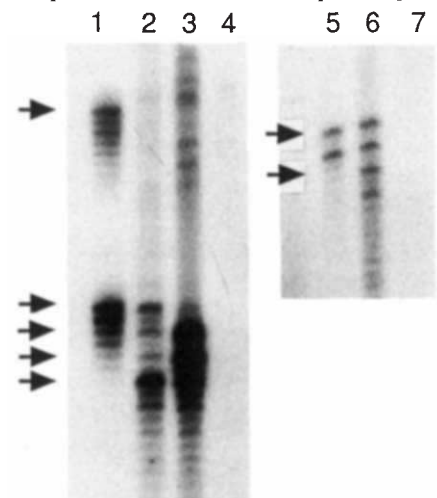
David Jones

DNA typing of museum birds

SIR — The recent report by Hagelberg *et al.*¹ of identification of human skeletal remains by DNA typing illustrates the synergistic effects derived from a combined use of the polymerase chain reaction (PCR) and microsatellite markers. While PCR greatly facilitates studies of ancient DNA (see Sykes's recent News and Views article²), microsatellites are highly feasible as genetic markers because of their abundancy and hypervariability³.

An interesting possibility in this context is to DNA-type museum specimens with the aid of PCR-analysed microsatellites. This would allow detailed genetic analyses of extinct species⁴ as well as population-genetic comparisons of present and past populations⁵. I have evaluated feathers as genomic DNA source for avian microsatellite analysis of museum specimens. Using microsatellite sequences isolated from the swallow (*Hirundo rustica*) and the pied flycatcher (*Ficedula hypoleuca*), I have typed birds of these species which are more than 100 years old.

DNA was prepared from the root part of single feather shafts (remiges and retrices provided by G. Frisk and S. Mathiasson, Swedish and Gothenburg Museums of Natural History) by a simple protease/boiling procedure in the presence of a chelating resin. Approximately 5% of a feather preparation was subjected to PCR, yielding autoradiography signals corresponding to 10–50 ng template DNA. None of the primer pairs



Examples of PCR amplifications of microsatellites from museum specimens of birds, separated by polyacrylamide gel electrophoresis. Lanes 1–4, amplifications from the pied flycatcher locus PTC3; lanes 5–7, amplifications from the swallow locus STG4. The birds were collected in 1924 (lane 2), 1926 (lane 3), 1890 (lane 5) and 1915 (lane 6). Lane 1 is a bird sampled in 1991, lanes 4 and 7 are negative controls for each locus, respectively. Alleles are arrowed.

amplifies modern human DNA so it is unlikely that the PCR products arise from human contamination.

An analysis of five pied flycatchers collected in Sweden between 1912 and 1926 reveals that all the birds were heterozygous, together showing five different 130–140-base-pair (bp) alleles at PTC3, a thymine–cytosine dinucleotide repeat. Three swallows collected between 1860 and 1930 showed three different alleles at STG1 and STG4, two thymine–guanine dinucleotide repeats (about 210 and 140 bp, respectively). All alleles of the museum specimens are present in extant populations in Sweden.

Most previous studies of ancient DNA have focused on multicopy mitochondrial DNA sequences. Although genomic microsatellites have been successfully amplified from skeletal remains¹, Hagelberg *et al.* warned of the low yields of nuclear DNA obtainable in bone preparations. Because old feathers seem to be

a good source for nuclear DNA, birds may be particularly well suited for DNA typing of museum specimens. The regular presence of large skin collections facilitates such approach.

A potential problem when typing old microsatellites may be numerous artefact bands generated through jumping PCR if DNA is degraded (ref. 5; S. Pääbo, personal communication). By comparing the ancient amplification products and the extra band-pattern obtained with modern DNA, discrepancies can be monitored.

HANS ELLEGREN

Department of Animal Breeding and Genetics,
Swedish University of Agricultural Sciences,
Box 7055, S-750 07 Uppsala, Sweden

- Hagelberg, E., Gray, I. C. & Jeffreys, A. J. *Nature* **352**, 427–429 (1991).
- Sykes, B. *Nature* **352**, 381–382 (1991).
- Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44**, 388–396 (1989).
- Thomas, R. H. *et al. Nature* **340**, 465–467 (1989).
- Pääbo, S., Irwin, D. M. & Wilson, A. C. *J. biol. Chem.* **265**, 4718–4721 (1990).

Symmetry and chaotic data

SIR — Lorenz¹ has reported numerical experiments that cast doubt on the validity of estimates of the correlation dimension derived from experimental data. His warnings about the effect of strong and weak coupling should be heeded. But one of the features of his model, symmetry, may be partially responsible for some of the difficulties observed. Because many parameters have been set equal in his numerical experiments, the equations are symmetric under certain permutations of the variables. Indeed, the symmetry group is abstractly isomorphic to that of a square ($4m$ in crystallographic notation).

The theoretical basis of all 'delay coordinate' reconstruction methods is the embedding theorem of Takens², which states that for 'generic' dynamical systems and observations, a suitable time-delay map is a topological embedding of phase space. However, symmetric systems are *nongeneric*, and violate the hypotheses of the Takens embedding theorem. For example, for reconstruction to work properly, certain eigenvalues must be simple — symmetry, however, can force them to be multiple. There is thus no theoretical basis for applying traditional methods of phase space reconstruction to symmetric systems. The reconstruction map may not only fail to embed phase space: it may even fail to immerse it (embed with self-intersections) and thus create singularities. These in turn can produce numerical instability, and invalidate standard procedures that require local linear approximations (such as box-

counts for dimensions). Instead, a symmetrically related set of observations must be used^{3,4}. Moreover, they must take values in a representation V of the symmetry group that is sufficiently complicated: the phase space must be subordinate to V in the sense of Wassermann⁵.

Lorenz contrasts dimension estimates and reconstructions performed using two distinct variables, Z_7 and Z^* . The variable Z_7 is asymmetric, and so is the attractor in Lorenz's Fig. 4a. However, Z^* is invariant under the full symmetry group, and phase space is not subordinate to the corresponding representation. It is not surprising that markedly different results are obtained.

It is therefore important to disentangle symmetry effects from those of weak versus strong coupling. This might be done by varying the coupling constants in Lorenz's model to destroy the symmetry, and repeating the analysis. When doing so it should be borne in mind that approximate symmetry will in practice lead to similar problems to those created by exact symmetry.

IAN STEWART

Mathematics Institute,
University of Warwick,
Coventry CV4 7AL, UK

- Lorenz, E. *Nature* **353**, 241 (1991).
- Takens, F. in *Dynamical Systems and Turbulence* (eds Rand, D. A. & Young, L. S.) 366 (Springer, New York, 1981).
- King, G. P. & Stewart, I. N. in *Nonlinear Equations in the Applied Sciences* (eds Ames, W. F. & Rogers, C. F.) 257 (Academic, Boston, 1991).
- King, G. P. & Stewart, I. N. *Phase Space Reconstruction for Symmetric Dynamical Systems* Preprint (Math. Inst., Univ. Warwick, 1991).
- Wassermann, A. *Topology* **8**, 127 (1969).

Keeping bloody relics liquid

SIR — L. Garlaschelli *et al.* (*Nature* **353**, 507; 1991) suggest an ingenious explanation for the periodic liquefaction of relics of saintly clotted blood. Their thixotropic mixture containing iron and calcium ions in solution would also provide an effective means for maintaining iron in the haem molecule at alkaline pH, particularly useful where the globin component had been long degraded (miraculous preservation excepting). But an even simpler recipe to ensure liquefaction would be to mix a little of the holy man's blood in one of the thixotropic honeys, of which heather (*Calluna vulgaris* (L.) Hull) would be one widespread and long-known example to be followed on appropriate occasions by a judicious, if not deific, shake *in camera*.

GEORGE A. F. HENDRY

ANTHONY J. E. LYON

Department of Animal and
Plant Sciences,
The University, Sheffield S10 2TN, UK

SIR — Does the thixotropic mixture manufactured by L. Garlaschelli *et al.* (*Nature* **353**, 507; 1991) meet all three criteria for the 'miracle'? First, frothing occurs following liquefaction; second, liquefaction takes up to 30 minutes and frequently longer to occur; and third, liquefaction occurs without the necessity of either tilting the container or turning it upside down. During the religious ceremony described by Garlaschelli *et al.*, the 'blood' is expected to liquefy, then 'boil' and become a frothy liquid. This procedure usually occurs after 40 minutes of fervent praying by a packed congregation. Should the liquefaction be sluggish, there is a great wailing of imprecations and even more fervent praying. Reputedly in the years when liquefaction has not occurred, there have been catastrophes — 1527, plague with more than 40,000 deaths; 1836, cholera with 24,000 deaths; 1980, an earthquake with more than 3,000 deaths.

An alternative explanation for the 'liquefaction' is that the reliquaries contain a mixture of spermaceti, ether and red dye from roots of the plant alkanet. This mixture is solid at 50 °F (10 °C), melts when held in the hand from body heat and, if any ether is present, froths and boils. The red dye that is released when roots of alkanet are crushed was used by the Romans to dye wool. Other compounds have been postulated in the past, most suggestions being a mixture of waxes and a red dye, but they failed to meet all three criteria.

JAMES M. DUNLOP

Hull Health Authority, Victoria House,
Park Street, Hull HU2 8TD, UK

Let them eat soap

SIR — Goodfellow and Sefton¹ have provided a glowing description of the progress in determining the structure of the human genome. While reading their News and Views article, and being dazzled by the recent identification and cloning of medically relevant genes, I could not help being pulled up short by their parenthetical comment (in an explanation of Kallman's syndrome, the gene for which has apparently been cloned) that in order to determine if an individual is unable to smell (anosmic), we are reduced to "asking patients to eat soap".

I would like to reassure readers, particularly any who feel they may have an olfactory deficit but balk at munching a bar of soap, that this is not the state of the art in chemosensory measurement. Although technical advances in olfactory psychophysics may not be as rapid or revolutionary as those in molecular genetics, the field has made progress in the refinement of measurement techniques. Quite frankly, I have never asked a patient to eat soap, nor am I aware of 'soap-eating' being advocated as a test for anosmia by anyone currently working in the area of clinical disorders of smell. There are, however, evaluative procedures that enable us to characterize both the nature and degree of chemosensory dysfunctions²⁻⁴.

BEVERLY J. COWART

Department of Otolaryngology,
Monell-Jefferson Taste and Smell
Center,
909 Walnut Street, Philadelphia,
Pennsylvania 19107, USA

1. Goodfellow, P. N. & Sefton, L. *Nature* **353**, 117-118 (1991).

2. Cain, W. S. *et al.* *Laryngoscope* **98**, 83-88 (1988).

3. Cowart, B. J. *Ann. N.Y. Acad. Sci.* **561**, 39-55 (1989).

4. Deems, D. A. *et al.* *Arch. Otolaryngol. Head Neck Surg.* **117**, 519-528 (1991).

Trivial error

SIR — Shapiro demonstrates¹, from data collected on DNA profiles at the FBI laboratory, that band-weight measurement errors within the same profile are correlated; and there are systematic differences between the operators who made the measurements.

It is difficult to understand why Shapiro should have been surprised by a correlation phenomenon that has been well documented elsewhere. In 1989, I and colleagues² said that "... errors in band positions within a DNA profile are highly correlated ...". P. Gill and I have shown how measurement-error correlations can be built into highly effective statistical procedures^{3,4}. We used a correlation coefficient of 0.9 — higher than all but one of those reported by Shapiro.

The second observation is not particularly surprising and it is difficult to see why Shapiro made so much of it. DNA profiling is so highly discriminating that in the event of a match the accuracy with which one estimates the population frequency is of minor importance. In our laboratory we have demonstrated, in experiments which have been reported at several international meetings and in the literature³, that even when we use deliberately distorted or unrepresentative databases the consequences within a forensic context are almost trivial.

Shapiro's points one to four are specious and will create unnecessary fears in the forensic science community, as our own experiments have shown. If I have any criticism of the methods of the FBI it is that its forensic scientists are unnecessarily cautious in the estimates of evidential strength that they derive from their frequency tables.

IAN W. EVETT

The Forensic Science Service,
Central Research and Support
Establishment,
Aldermaston, Reading,
Berkshire RG7 4PN, UK

1. Shapiro, M.M. *Nature* **353**, 121-122 (1991).

2. Evett, I.W. *et al.* *Nature* **340**, 435 (1989).

3. Evett, I.W. & Gill, P. *Electrophoresis* **12**, 226-230 (1991).

4. Evett, I.W. *et al.* *Int. Symp. Hum. Identification 1989* 77-101 (Promega, Madison, Wisconsin, 1990).

Neutral mutation hypothesis test

SIR — McDonald and Kreitman¹ claim that adaptive mutations are largely responsible for the evolution of alcohol dehydrogenase (Adh) because, according to their calculations, in the Adh gene the ratio of nonsynonymous to synonymous substitutions between three *Drosophila* species (7:17) is much larger than the ratio (2:42) within species. However, their test has at least the following problems.

First, McDonald and Kreitman neglect the between-species variation at a nucleotide site, if that site is polymorphic in any of the three species. Both this counting rule and the one we discuss below tend to underestimate between-species variation. For example, at position 1,590 all the alleles in *D. melanogaster* have T, half the alleles in *D. simulans* have T and the other half have C, and all the alleles in *D. yakuba* have C. At this position, they assigned a single polymorphic synonymous substitution within *D. simulans* but no substitution between species. However, under their assumption of monomorphism in the ancestral species, an additional synonymous substitution must have occurred between *D. melanogaster* and *D. yakuba*.

Second, they fail to consider the fact

REPLACEMENT (NONSYNONYMOUS) AND SYNONYMOUS DIFFERENCES IN THE CODING REGION OF THE ADH GENE

Type of difference	Within species			Between species*		
	<i>mel</i>	<i>sim</i>	<i>yak</i>	<i>mel</i> vs <i>sim</i>	<i>mel</i> vs <i>yak</i>	<i>sim</i> vs <i>yak</i>
Replacement (<i>R</i>)	0.71	0.00	0.00	2.58 (2.22)	6.58 (6.22)	6.00 (6.00)
Synonymous (<i>S</i>)	5.54	4.60	3.16	10.67 (5.60)	28.10 (23.75)	24.29 (20.41)
<i>R/S</i> ratio	0.13	0.00	0.00	0.24 (0.40)	0.23 (0.26)	0.25 (0.29)

*Numbers in parentheses denote between-species differences after subtraction of within-species differences.

mel, *D. melanogaster*; *sim*, *D. simulans*; *yak*, *D. yakuba*. (The 12 alleles in *D. yakuba* in Table 2 of ref. 1 probably represent 24 alleles, and this has been assumed in our analysis.)

that a polymorphism in a species does not necessarily imply that the polymorphism arose within that species. In particular, because *D. simulans* and *D. melanogaster* are closely related sibling species, a polymorphism in one species may also indicate that the ancestral species was polymorphic, and that polymorphism was subsequently lost in one of the descendent species.

Third, their decision as to whether or not a site is polymorphic is highly sensitive to the number (*n*) of sequences sampled and *n* is only 6 for *D. simulans*, yet their test neglects sampling errors. Thus, even if one ignores all other problems, their test is still not acceptable. If the numbers of fixed and polymorphic nonsynonymous substitutions are 6 and 3 instead of 7 and 2, then Fisher's exact test gives a probability of 7% (nonsignificant). Thus, the conclusion is sensitive to sampling errors.

A simple way to overcome these problems is as follows. First, compute the number of substitutions per synonymous site (K_s) and per nonsynonymous site (K_a) between every pair of sequences, both within- and between species. Second, for each species, compute the average within-species K_a and K_s values and the K_a/K_s ratio. Third, compute the average between-species K_a , K_s and K_a/K_s values for every pair of species. Fourth, compare the between-species K_a/K_s ratios with the within-species ratios. In the present case, however, the sequences from *D. yakuba* were obtained by directly sequencing the polymerase chain reaction products, so that the two alleles in an individual at the ADH locus cannot be clearly defined when they differ by more than one nucleotide. For this reason, we use the following slightly less rigorous method. We compute the numbers of synonymous and nonsynonymous nucleotide differences (*S* and *R*, respectively) between every pair of sequences at each of the sites listed in Table 1 of ref. 1. For instance, consider position 816. In *D.*

melanogaster, five alleles have T and seven have G. Because the difference between T and G at this position is synonymous, $R = 0$ for every sequence pair, and the average *S* value is $35/66 = 0.53$, because $5 \times 7 = 35$ out of $12 \times 11/2 = 66$ possible pairwise comparisons have the $G \leftrightarrow T$ difference. In *D. simulans*, all six alleles have T. Thus, between *D. melanogaster* and *D. simulans*, $R = 0$ and the average *S* is $7 \times 6/12 \times 6 = 0.58$. The *R* and *S* values are then summed over all sites (see table).

From the total *R* and *S* values, we find that the *R/S* ratios between species are remarkably similar among the three comparisons: 0.24, 0.23 and 0.25, the average being 0.24. If we assume that the ratio found in the between-species comparisons also holds for the within-species comparisons, then the expected within-species *R* values are 1.33, 1.10, and 0.74 for *D. melanogaster*, *D. simulans*, and *D. yakuba*, respectively. The differences between the expected values and the observed ones (0.71, 0 and 0) are not large enough to be statistically significant, for the variance of *R* is given by $R(n+1)/[3(n-1)] + 2R^2(n^2+n+3)/[9n(n-1)]$ (ref. 3).

In the table we also compute the net between-species differences using the formula $d = d_{xy} - a(d_x + d_y)/2$, where d_{xy} denotes the differences between species X and Y, and d_x denotes the within-species differences in species X. Note that both the synonymous and replacement differences between *D. yakuba* and either of the other two species are much larger than the within-species values; for example, $S > 20.0$ versus $S < 6.0$. This is in sharp contrast to McDonald and Kreitman's calculation that only 17 synonymous substitutions had occurred between species whereas 42 had occurred within species. For the net between-species differences the *R/S* ratios are now 0.40, 0.26 and 0.29 with an average of 0.32 (we take the simple average, though the last two ratios are more favourable for our argument, and would

be more reliable, for they are based on larger *R* and *S* values). With this ratio, the expected within-species *R* values are 1.77, 1.47 and 1.01 for *D. melanogaster*, *D. simulans*, and *D. yakuba*, respectively. The total expected *R* value (4.25) for the three species differs from the total observed value (0.71) by 3.56, which is smaller than 1.96 times the standard error 1.86, which is an underestimate, for it neglects the positive evolutionary correlations between *R*s and the errors in estimating the expected *R/S* ratio and the within-species *S* value. Therefore, the neutral hypothesis cannot be rejected, even under unfavourable assumptions.

A more powerful test is to use Watterson's theory⁴. For example, for *D. melanogaster*, $n = 12$ and $k = 14$ (number of segregating sites in the sample) for synonymous sites and so from formula (1.4a) of Watterson we obtain $\theta_s = 4.64$. If $R/S = 0.32$ $\theta_R = 1.48$ for replacement (nonsynonymous) mutations and from formulas (1.4a) and (1.4b) we obtain the mean and variance of *k* as 4.48 and 7.91 for nonsynonymous sites. The corresponding mean and variance are 3.52 and 7.00 for *D. simulans* and 5.76 and 9.57 for *D. yakuba*. The expected *k* values are not significantly different from the observed values (2, 0 and 0). If we assume that the three species are independent and pool the data together, then the total *k* is 13.76 ± 4.95 . Taken at face value, this expected *k* differs significantly from the observed value 2 (at the 2% level). However, the significance is uncertain because the estimate of s.e. neglects all the errors mentioned in the preceding paragraph. If $R/S = 0.24$, then *k* becomes 10.3 ± 4.04 , which is not significantly different from the observed value if the above-mentioned errors are included in the estimation of s.e.

In conclusion, it is not clear as to whether the ADH data can be taken as evidence against the neutral hypothesis.

DAN GRAUR

Department of Zoology,
Tel Aviv University,
Ramat Aviv 69978, Israel

WEN-HSIUNG LI*

Center for Demographic and
Population Genetics,
University of Texas,
PO Box 20334, Houston,
Texas 77225, USA

* To whom correspondence should be addressed.

SIR - Comparing nucleotide sequences of the alcohol dehydrogenase (Adh) gene within and between three species of *Drosophila*, McDonald and Kreitman¹ concluded that the number of non-synonymous (amino-acid altering) nucleotide substitutions is significantly greater than expected under neutral pro-

cesses and that the excess substitutions result from fixation of selectively advantageous mutations. This conclusion is based on a statistical test of the prediction that under neutrality "the ratio of replacement [nonsynonymous] to synonymous fixed substitutions should be the same as the ratio of replacement to synonymous polymorphisms." We believe that there are subtle but serious problems in McDonald and Kreitman's reasoning.

The prediction as derived from the neutral theory can be more precisely stated as a null hypothesis as follows: for strictly neutral mutations, the ratio of nonsynonymous to synonymous substitutions between species is equal to the ratio of nonsynonymous to synonymous substitutions between alleles within species. A proper test of this hypothesis requires estimates of the number of nucleotide substitutions per site that have occurred during the divergence of alleles within species and between species. However, the *G*-test used by McDonald and Kreitman does not compare rates of nucleotide substitution but instead is based on the ratio of synonymous to nonsynonymous sites that have been classified as polymorphic within or fixed between species. Such a classification of sites is arbitrary because fixation (or polymorphism) at a site in a population is a transient event in evolution. In addition, the number of polymorphic and fixed sites depends, to an unknown extent, on both number of sequences examined as well as the number of species studied. This second point is illustrated, for example, by site 816 in Table 1 of ref. 1, which would have been a fixed site if only *D. simulans* and *D. yakuba* were studied, but was classified as a polymorphic site because it was polymorphic in *D. melanogaster*. Although these effects may have only minor influence on the ratio of nonsynonymous to synonymous sites that are polymorphic or fixed in a sample, there is virtually no mathematical theory for predicting the distribution of the number of polymorphic and fixed sites across species under the neutral hypothesis. (Hey recently studied⁵ the sampling distribution of fixed differences between two species.) Finally, since both the number of fixed and polymorphic sites are subject to large stochastic errors^{4,5}, the stochastic variances of these quantities should be taken into account in a valid test of the difference between the two ratios.

We suggest that as a general test of the null hypothesis of equal ratios one should evaluate the average nucleotide substitutions for all pairwise comparisons of sequences within and between species (π and d_A , respectively)⁶. Our test uses the frequency of nucleotides in the sam-

ple, accounts for multiple substitutions at sites, and incorporates stochastic errors in the evolutionary process. We first reanalysed McDonald and Kreitman's data for *D. melanogaster* and *D. yakuba*. We estimated d_A and the average π for the two species for both synonymous and nonsynonymous sites using the Nei-Gojobori method⁶. The standard errors of these quantities were computed by taking into account stochastic variances^{6,7}. The results obtained were $d_A = 14.20 \pm 3.71$ and $\pi = 2.40 \pm 1.01$ per 100 synonymous sites and $d_A = 1.1 \pm 0.44$ and $\pi = 0.06 \pm 0.05$ per 100 nonsynonymous sites. Therefore, the ratio of nonsynonymous to synonymous substitutions is 0.077 ± 0.037 between species and 0.026 ± 0.024 within species. Although the former ratio is higher than the latter, the difference is not statistically significant ($Z = 1.2$; $P > 0.2$). (The smaller ratio within species could be partly due to excess synonymous substitutions caused by the balanced polymorphism of the *F* and *S* alleles in *D. melanogaster*⁸.) The analysis for three species, which is somewhat more complicated, gives essentially the same results. Thus, these results do not support the conclusion that there is a significant excess of nonsynonymous substitutions resulting from adaptive fixation of mutations.

THOMAS S. WHITAM
MASATOSHI NEI

*Institute of Molecular Evolutionary
Genetics,
Department of Biology,
Pennsylvania State University
University Park,
Pennsylvania 16802, USA*

MCDONALD AND KREITMAN REPLY — Graur and Li, and Whittam and Nei, point out that a polymorphism in one species may have arisen in an ancestral species; that different polymorphic sites have different allele frequencies; and that the number of polymorphisms will increase as more alleles and more species are sampled. For the neutral model of molecular evolution that we tested, all of these phenomena will affect neutral replacement substitutions and neutral synonymous substitutions equally. Therefore they do not affect the validity of our test of the neutral model, and they are not alternatives to adaptation as explanations for the ratio of replacement to synonymous substitutions being much greater for fixed differences than for polymorphisms at the *Adh* locus in three species of *Drosophila*¹.

The authors also question our method for counting a site which has one nucleotide fixed in one species, a different nucleotide fixed in a second species, and both nucleotides polymorphic in a third species. They suggest that we count such

a site as one fixed difference and one polymorphism, rather than just as one polymorphism. Any rule for classifying substitutions as fixed or polymorphic will affect neutral replacement and neutral synonymous substitutions equally, and it is only important to apply the same rule to both. We choose to count as fixed substitutions only those that are fixed in every species in which they appear. This is because under the alternative hypothesis of adaptive fixation of replacement mutations, a replacement substitution that is adaptive in one species, and thus rapidly becomes fixed, will either be adaptive or maladaptive in other species, and thus is unlikely to be polymorphic in any species. There is also a practical reason for counting each substitution only once, rather than trying to estimate the number of times that the substitution has gone to fixation. Estimating the number of fixation events would require an accurate estimate of the species' phylogeny; we think it is an advantage of our test that it requires no such estimate.

The authors of the above letters suggest tests based on the gene diversity between and within species, rather than the numbers of fixed and polymorphic substitutions. These tests use estimates of the stochastic variance, which is the variation in gene diversity among loci resulting from the different coalescent times (times to the most recent common ancestor) of different loci^{3,7}. For a single locus, however, replacement substitutions are intermingled with synonymous substitutions, and thus replacement and synonymous substitutions have the same coalescent time. Therefore only sampling variance needs to be considered. It would be interesting to see the results of a test using gene diversity that used the sampling variance, which is much smaller than the stochastic variance. Whatever the outcome of such a diversity-based test, however, our substitution-counting test remains valid. We suspect that any diversity-based test will be more complicated, will require more assumptions, and will be less statistically powerful than our method for detecting adaptive protein evolution.

JOHN H. MCDONALD
MARTIN KREITMAN

*Department of Ecology and
Evolutionary Biology,
Princeton University,
Princeton, New Jersey 08544, USA*

1. McDonald, J. H. & Kreitman, M. *Nature* **351**, 652–654 (1991).
2. Li, W.-H., Wu, C.-I. & Luo, C.-C. *Molec. Biol. Evol.* **2**, 150–174 (1985).
3. Tajima, F. *Genetics* **105**, 437–460 (1983).
4. Watterson, G. A. *Theor. Pop. Biol.* **7**, 256–276 (1975).
5. Hey, J. *Genetics* **128**, 831–840 (1991).
6. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
7. Takahata, N. & Nei, M. *Genetics* **110**, 325–344 (1985).
8. Kreitman, M. in *Evolution at the Molecular Level* (eds Selander, R. K., Clark, A. G. & Whittam, T. S.) 204–221 (Sinauer, Sunderland, Massachusetts, 1991).

Grasping nettles

Jeremy Purseglove

Wetlands: A Threatened Landscape. Edited by Michael Williams. *Blackwell: 1991.* Pp.419. £70, \$79.95.

THIS is a very dry book on wetlands. It is, nonetheless, a great mine of information, assembled by a distinguished group of authors, and provides a solid background to the wetland debate. Wetlands are guaranteed to fuel controversy if, as recognized in the last chapter, their conservation involves "redistributing of rights to use the environment and by implication an allocation of wealth". Legislating to resolve these debates is, as the authors frequently suggest, fraught with the difficulty of conflicting subjective values. Still, if their creed that wetland destruction "should not be undertaken without much thought and good cause" is not to be totally flouted, some nettles will need to be grasped. In the developed world an over-powerful and greedy agricultural lobby should not be able to lay its hands on inappropriate subsidies for wetland drainage. And those who grant foreign aid to developing countries need to understand that local people are often best served economically by harvesting indigenous wetlands, or at least by partial reclamation and mixed cropping, rather than by grandiose imported drainage schemes.

"The most comprehensive appraisal of the world's wetlands ever published", is the claim on the dust jacket. The main strength of the book certainly is its breadth. Eutrophication and tourism are described on Lake Dal and on the Norfolk Broads. The port expansion at Singapore can be compared here with that at Rotterdam. There are accounts of mediaeval drainage in the Netherlands, reclamation of the Sundarbans by the British Raj and the impact of the great Gold Rush on the wetlands of San Francisco Bay. There are especially good sections on tropical wetlands.

In spite of its wealth of sources, I was always hungry for yet more examples rather than generalizations. The difficulties of conserving the potholes of North Dakota, the plight of the Florida panther and the spectacular wetland creation schemes of the US shooting organization Ducks Unlimited were typical details constantly needed to bring the issues to

life. The archaeological chapter is an exception to the rest in being vivid with detail. Wetlands near Trento have preserved a cloche hat of woven reed and guelder rose worn by some chic prehistoric Italian. In South Australia, 10,000-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Swamped by possibilities — land ripe for conservation, preservation, commercialization and deforestation.

year-old boomerangs were found in a wetland that still supports the species of tree from which the boomerangs were made. Such images really bring home the value of an ancient wetland and are worth pages of academic jargon, here epitomized by a fatuous three-page table describing different recreational attitudes to wetlands ranging from the "negativistic" (killing and draining) to the "moralistic" (animal rights).

A more understandable problem, presumably because of the ambitious breadth of the book, is the lack of space for deep exploration of specific issues. Thus peatlands are touched on, but there is no real discussion of the dilemmas and arguments surrounding afforestation of the Flow Country in Scotland, or the excavation of lowland raised mires in the United Kingdom and the Soviet Union for horticultural peat. Similarly, the increasingly urgent issue of water abstraction receives little coverage.

Where specific issues are addressed, some conclusions do emerge. As a destroyer of wetlands, recreation is not in the same league as urban development or intensive agriculture. Although tourists at Kakadu National Park in Northern Australia "couldn't be trusted with rifles" and even "got themselves into trouble by tempting the crocodiles" (one solution to the problem?), careful design of campsites and access roads minimized the tourists' impact on the reserve. For similar reasons, the effect of 400,000 annual visitors to the Everglades is reported as "slight". The big business in tourism, such as that on the shores of Lake Erie, may even prove to be the salvation of wetlands by warding off other threats.

Planet Earth Pictures

The case is well made that, for the wetlands of developing countries, both human inhabitants and wildlife are worse off after major drainage schemes. Thus drainage of the Sudd could leave most of those without access to land even having to import their own food. But the authors are also honest enough to accept that swamps and forests converted to rice paddy-fields in South-East Asia are now "immensely productive wetlands which feed great numbers of human beings every year". Wholesale generalizations can no more be applied to the wetland debate any more than to any other issue.

Maps are generally excellent. Photographs vary from a splendid shot of buses marooned in a Somerset flood to what appears to be the professor's holiday snaps taken with his brownie box. Blobs in the distance may be hippos, storks or mud skippers, but you will need to check the caption to be sure. Publishers have no excuse for wasting space in a £70 book on such pictures, as if a work of scholarship is somehow demeaned by attractive presentation. Nevertheless, for those who can afford it, this is a remarkably wide-ranging compendium of knowledge about wetlands. □

Jeremy Purseglove is a landscape architect at the civil engineering firm Mott MacDonald, Cambridge, UK, and author of Taming the Flood, which describes the history and current status of the rivers and wetlands of Europe (Oxford University Press, 1989).

■ *Wetlands* has just been published by Facts on File, in association with The International Waterfowl and Wetlands Research Bureau (general editors, Max Finlayson and Michael Moser). Price is £19.95. □

Seeing in the dark

Philip Kitcher

Scientism: Philosophy and the Infatuation with Science. By Tom Sorell. Routledge: 1991. Pp. 206. £35, \$45.

ABOUT 20 years ago, in an informal seminar held in connection with a drama festival at the University of Cambridge, one of the most prominent literary critics of our time suggested that Molière and Stendhal have more to teach us about the workings of the human mind than any conglomeration of academic psychologists. Whether or not the remark was intended as a serious thesis about the limitations of science, it has remained with me as an exemplary formulation of the concerns that underlie accusations of scientism. At a time when various intellectuals are again offering more-or-less strident denunciations of the alleged hegemony of natural science, it is salutary to focus the issue on a clear and interesting claim: is the psychological understanding achieved by the great dramatists, poets and novelists superior to that gained in the laboratory?

Despite its title and some of the comments on the back jacket, Sorell's *Scientism* is a measured book. Sorell believes that much scientific work is valuable, and worries only that our appreciation of its qualities will blind us to the virtues of nonscientific approaches. Indeed, he is sufficiently concerned about some modes of denouncing science (for example the charges levelled by 'creation scientists') to limit his argument to a narrow domain. Philosophy, he suggests, has been overawed by the natural sciences. As a result, since the seventeenth century, generations of philosophers have adopted pale imitations of what they take to be scientific thinking in an attempt to resolve philosophical problems. Unfortunately, the really deep problems of philosophy resist such approaches, and, cutting themselves off from fruitful, nonscientific sources of ideas, philosophers doom themselves to failure.

Sorell's strategy for defending this position is to chip away on many different fronts. He begins with those rather predictable targets, the logical positivists, apparently everyone's favorite villains these days. (It is worth recalling not only the intellectual liberation brought about by the Vienna Circle, but also the steadfast efforts of their members in opposing Fascism.) Turning his attention next to the seventeenth century, he tries to expose the roots of philosophy's infatuation with science, arguing that both

Bacon and Descartes slighted the credentials of nonscientific branches of learning.

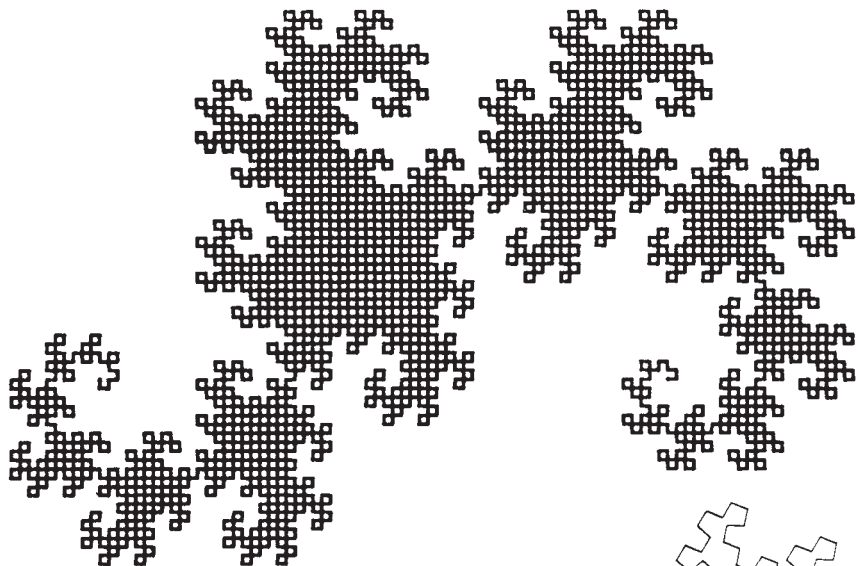
Sorell continues with some fairly detailed textual exegesis — and, to my mind, a certain amount of special pleading — designed to show that Kant was on the side of the righteous. Despite his obvious respect for eighteenth-century science, Kant recognized the value of other modes of thinking, notably our moral reasoning. But instead of trying to develop a kantian conception of knowledge and action, Sorell changes focus, offering first some sensible reflections on moral assessments of science, then an ecumenical look at the 'two cultures' debate, and finally some specific criticisms of 'scientistic' ventures in contemporary Anglo-American philosophy. These pages are dotted with useful points, but, in my judgement, they will leave the reader unsatisfied.

By the end of the book, we know that Sorell respects the employment of science in its proper sphere. We also know that he believes that there are nonscientific modes of generating insight, pre-eminently in the realms of moral thinking and in the arts. What is lacking is any systematic treatment of exactly what kinds of problems resist assimilation to the ideas or methods of the natural

sciences, and of exactly what kinds of ideas or methods might be appropriate to these problems. For all its piecemeal criticism, Sorell's book never provides a crisp thesis about the rival claims of two well-defined modes of inquiry — as, for example, the suggestion that the great dramatists have more to teach us about the springs of human action than do the academic psychologists.

The absence of a definite positive view weakens some of Sorell's criticism. In one of his later chapters, he takes on the suggestion, offered by Patricia Smith Churchland, that contemporary developments in neuroscience can reform the theory of knowledge and the philosophy of mind. Sorell believes that the really big problems about knowledge and mind will always elude reductionist analyses. Perhaps. But just as philosophy has played midwife to the sciences, the sciences have tutored philosophy. The 'new science' of the seventeenth century advanced philosophical discussions in epistemology and metaphysics. Developments in nineteenth-century mathematics have transformed logic. Recalling an earlier period, we can imagine a fourth-century-BC Tom Sorell protesting that the really big issues about change and motion will always resist scientific solution, not foreseeing the ways in which mathematics

Images of chaos



IMAGINE folding a narrow strip of paper in two, and then repeating this twice, folding in the same direction. On unfolding, you would have a strip that from the side would appear as a meandering line, the vertices of which are folds. With a computer, you could draw the meandering line that results from folding any number of times. The line on the right would result from six folding steps. The 'dragon' curve shown above is a special case resulting from 14 steps and is so called because it reminded its discoverer, J. E. Heighway, of Chinese dragons. Surprisingly, it does not intersect itself and has the same basic form as the line below. In *Fractals*, Hans Lauwerier introduces the concept of fractals by discussing these and other meandering lines, making clear the underlying mathematics of fractals as well as explaining how to conjure up these "endlessly repeated geometric figures" on computer screens. Accessible and colourful, the book is translated by S. Gill-Hoffstadt and published by Penguin/Princeton University Press. Price £9.99, \$14.95.

and physics have resolved Zeno's deep paradoxes.

Drawing on these parallels, Churchland can surely issue a counter-challenge. After two millennia of wrestling with conundrums about knowledge and mind, during which advances have been made by relying on and contributing to the nascent sciences, philosophy should continue to look to what sources it can for possible lines of solution. Neuroscience is plainly a possible guide. Claiming that it cannot help us to solve the really big problems is a hollow rejoinder unless we are offered some alternative place in which to seek insight.

Balanced views?

Brian Charlesworth

Fifty Years of Genetic Load: An Odyssey. By Bruce Wallace. Cornell University Press: 1991. Pp.174. \$31.50.

BRUCE Wallace has had a long and distinguished career as an experimental population geneticist. Together with Theodosius Dobzhansky, he was one of the founders of the 'balance' theory of variation in natural populations. This theory was elaborated in the early 1950s as an alternative to the 'classical' view, whose chief exponent was the geneticist H. J. Muller, famous for his work on the fruitfly *Drosophila*. The classical view holds that nearly all loci in the genome are predominantly represented in a population by a wild-type allele, and that the rare variant alleles at a locus are deleterious and maintained solely as a result of mutation pressure. Because there are many loci in the genome of a higher organism such as *Drosophila* or humans, the presence of these rare, deleterious alleles contributes to a substantial reduction in the average darwinian fitness of a population below that of a hypothetical, mutation-free individual with maximum fitness. This reduction in fitness, measured relative to the maximum fitness, is the genetic load of the population. In addition to mutation, genetic load may be created by genetic variation actively maintained by natural selection, for example by a fitness advantage to the heterozygote for a pair of alleles over the two homozygotes for the alleles.

According to the balance view, a substantial proportion of the loci in the genome exhibit allelic variation that is actively maintained by such 'balancing' selection. The first crucial test of the predictions of the two rival theories was provided in the 1960s by the detailed

Scientism owes its readers a vision of how the great philosophical questions could be fruitfully approached without making use of the deliverances of science. Perhaps we need a thoughtful exploration of just how much Stendhal and Molière have to teach us about ourselves, and just how their approaches lie beyond the reaches of science. Until then, it is premature to chide those whose philosophical explorations are illuminated by their knowledge of science. Darkness is no substitute. □

Philip Kitcher is in the Department of Philosophy, University of California, San Diego, La Jolla, California 92093, USA.

examination of the level of natural variability at samples of loci, by means of the gel electrophoresis of many different proteins. The classical theory apparently failed this test, because a pattern of widespread variability at many loci was usually detected in population surveys. But at about the same time, the neutral theory of molecular evolution and variation was being elaborated. According to this theory, random sampling of alleles in finite populations can interact with mutation to produce a pattern of genetic variability similar to that found in natural populations, provided that the fitness effects of the mutant alleles are sufficiently small. Thus, the mere observation of extensive genetic variation does not confirm the balance view: it has to be shown that the variation is related to fitness.

Although there are a few cases where the action of selection on protein variability has been unequivocally established, the difficulties of experimentally detecting changes in fitness are so great that the issue is so far undecided. The rapidly growing body of evidence from DNA population studies has already made it clear that there is usually much more evolution and variation for nucleotide changes that do not affect the amino-acid sequence of proteins than for changes altering protein sequences. This indicates that most amino-acid changes are deleterious, and that most variability at the DNA level probably fits the neutral model. The increasing sophistication of statistical techniques for detecting the footprints left by selection on DNA and protein sequences offers some hope that the role of selection in maintaining variability will soon become clearer.

Wallace's book provides a personal account of the course of his work and thought on the nature and significance of genetic load. He describes how he experimented to determine the effects of irradiation on *Drosophila* populations; to his surprise, and contrary to the predictions of mutational-load theory, he found no relation between radiation

treatment and measures of the fitness of flies in his populations, despite evidence for a high frequency of genes with deleterious homozygous effects on fitness. This led him to conclude that heterozygote advantage is commonly responsible for the maintenance of genes with deleterious homozygous fitness effects, a keystone of the balance theory. To test whether random gene changes often induce mutations with favourable heterozygous fitness effects, he did a series of ingenious and laborious experiments with irradiated *Drosophila* chromosomes. The experiments confirmed his suspicions. This extraordinary finding stimulated much research, resulting in a welter of conflicting results.

Wallace emphasizes the results that confirmed his ideas, and does not review all the relevant literature. The consensus view today is probably that nobody yet knows how to reconcile the conflicting results of the radiation experiments, and that they do not provide unequivocal support for the balance hypothesis. Nonetheless, Wallace holds firm to his belief in pervasive balancing selection, and devotes several chapters to expounding his views on how such selection, acting at many loci, is consistent with the survival of the population, despite the large genetic load that would be imposed if all loci acted independently. He plumps for a combination of density-dependent regulation of population size, and selection based on only the fittest proportion of the population contributing to the next generation. Again, the treatment is one-sided, and objections to his views are not adequately discussed.

It should be clear from these criticisms that Wallace's book cannot be recommended as an introduction to the current state of play on the topic of maintenance of genetic variation; nor is it a reliable history of this important aspect of evolutionary genetics. Rather, it is of interest to those who wish to understand how he came to arrive at his views, which are by no means widely accepted today but which have certainly been important in stimulating experiments on natural variation at the level of the gene. □

Brian Charlesworth is in the Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA.

New in paperback

■ Prentice Hall have recently published *The Edges of Science: Crossing the Boundary from Physics to Metaphysics* by Richard Morris. The book is a superb introduction to the deepest questions in physics and cosmology. Price is \$10.

■ In *Paper Heroes*, Witold Rybczynski gives a thought-provoking exploration of the relationship between technology and development. Published by Penguin, price \$9.95.

Spreading it out

R. H. Tredgold

An Introduction to Ultrathin Organic Films: From Langmuir–Blodgett to Self-Assembly. By Abraham Ulman. Academic: 1991. Pp.442. \$65, £46.

ORGANIC chemists have learned to make progressively larger and more complex molecules but are not so good at assembling such molecules in ordered ways to perform specific functions, as living things do. This is hardly surprising considering the difficulties involved. Polymer chemists have produced one-dimensional polymeric systems; even so, their monomeric precursors are usually quite small and simple. The assembly of ordered planar two-dimensional systems seems an obvious first step towards forming multimolecular functional systems. Over the past decade, there has been much work in this field, which this book aims comprehensively to review.

Of the various techniques available, the Langmuir–Blodgett method of forming ordered monolayers and multilayers has attracted most attention. Here a monolayer of insoluble amphiphilic molecules is spread on an aqueous sub-phase and is thence transferred to a solid substrate. Under suitable conditions, progressive raising and lowering of the substrate through the air–water interface leads to the deposition of a succession of bilayers on the substrate. The rather misleadingly named self-assembly method of forming ordered multilayers has also recently become popular. This technique has a marked similarity to the Merrifield method of peptide synthesis and involves immersing the substrate in a succession of different reagents so that a succession of monolayers are deposited on one another, linked vertically by covalent bonds. Ulman's own research is largely into self-assembly.

The first three parts of the book deal with diagnostic techniques, Langmuir–Blodgett films and films formed by self-assembly. In the fourth part, computer modelling of monolayers is discussed. So far, the approximations that have to be made render the predictions of modelling of dubious value. In the last section, several applications of ordered multilayers are discussed. This topic has already been overworked elsewhere and coverage here could well have been omitted.

The field of ultrathin organic films is now too large to be explored in both breadth and depth in a monograph of this kind. So Ulman has gone for breadth. Nonetheless, he has overlooked several important papers, even in his own field. But this is not a serious

criticism: it would be a superhuman task to provide a comprehensive list of relevant references for, say, the past ten years. The book will provide an excellent introduction to the subject and a valuable quarry for workers in the field anxious to widen their horizons. □

R. H. Tredgold is in the Department of Physics, University of Lancaster, Lancaster LA1 4YB, UK, and also in the Department of Chemistry, University of Manchester, Manchester M13 9PL, UK.

Burial laws

Douglas Palmer

The Processes of Fossilization. Edited by Stephen K. Donovan. Belhaven/Columbia University Press: 1991. Pp.303. £45, \$55.

THE *post-mortem* processes that act on cadavers can transform them beyond recognition and often remove all traces of their previous existence. It is no wonder that the general recognition of fossil petrifications as the remains of once living animals and plants did not happen until the late seventeenth century. Perhaps more surprising is that it has taken so long to understand the processes of fossilization. This book helps to explain the delay and demonstrates why multifactor processes such as fossilization are not easily reducible to models or formulae.

To begin with, there is the question of what sort of sample of the fossil record is available. With more than 1.5 million living species and assuming an average

species duration of 10 million years, over the more than 600 million years of life on Earth one would expect quite an extensive fossil record. Yet only about 150,000 fossilized species have so far been described. Just how representative this sample is has been a matter of concern from well before Darwin's time. The first third of the book reviews the problem and there is a brief attempt to put geological completeness into the broader context of information theory. It remains to be seen how much of the information loss really is redundant background noise.

The word 'taphonomy' was coined by I. A. Efremov in 1941 for the study of the 'laws' governing the transition of organic remains from the biosphere to the lithosphere. These 'laws of burial' are difficult to define. An important approach to the problem has been that of the German 'actualist' or neontological tradition of studying how organic remains are 'recruited' into the sedimentary record. The work goes back to Otto Abel at the beginning of this century but was not fully appreciated until the 1972 publication of the English edition (Oliver and Boyd) of W. Schäfer's *Aktuo-Paläontologie nach Studien in der Nordsee* (W. Kramer, 1962). Inevitably these studies were biased towards the classic sedimentary environments and biota of northern European seas. This kind of research has now been extended into other important sedimentary environments, an example being A. K. Behrensmeier's work on African vertebrate taphonomy.

K. M. Parsons and C. E. Brett review the concept of 'taphophacies' or groupings of assemblages based on similar taphonomic histories and its usefulness for interpreting ancient depositional environments. The next step is to examine the relationship between predicted death assemblages and the actual accumulated fossil record as seen in temporal sequences at specific localities. This will require a large and probably expensive programme of coring through from the contemporary sediment surface to below the so-called 'taphonomically active zone'. Only then can we finally cross the boundary between the actualist studies and those of fossil assemblages.

More than half of the book is devoted to the taphonomy of different fossil groups, from soft-bodied animals in general, through plants and foraminifera to vertebrates. All this is useful for those interested in this fascinating interface between neontology and palaeontology. 'Taphonomy' may be a bit young for my edition of the *Oxford English Dictionary*, but it is certainly here to stay. □

Douglas Palmer is at the National Museum of Wales, Cathays Park, Cardiff CF1 3NP, UK.



Leaves of the sumac tree, *Rhus stellariaefolia*, preserved in a shale, laid down in a freshwater lake some 40 to 50 million years ago. From the new edition of *Fossils: The Key to the Past* by R. Fortey, published by the British Museum (Natural History)/Harvard University Press. £12.95 (pbk), \$29.95 (hbk).

The galactic distribution of free electrons

J. M. Cordes*, J. M. Weisberg†, D. A. Frail‡, S. R. Spangler§ & M. Ryan||

* Astronomy Department and NAIC, Cornell University, Ithaca, New York 14853, USA

† Department of Physics and Astronomy, Carleton College, Northfield, Minnesota 55057, USA

‡ National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA

§ Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

|| Department of Physics, Cornell University, Ithaca, New York 14853, USA

Dispersion, distance and scattering measurements of pulsars and other radio sources are now sufficiently numerous to allow modelling of the entire galactic distribution of free electrons. A two-component axisymmetric model of local electron density, fitted on all scales from 100 km to a few parsecs, accounts for most of the data; a population of dense, discrete clouds is also needed, and there is some evidence for spiral structure.

DIFFUSE, ionized gas in the interstellar medium (ISM) is an important tracer of activity in the Galaxy that produces ionizing photons, such as radiation and stellar winds from massive stars and supernova blast waves¹. At present, the existence of ionized gas at large distances from the galactic plane is poorly understood and there are few direct constraints on the distribution of ionized gas in the inner Galaxy. Here we synthesize various radioastronomical measurements to derive a large-scale galactic model for the distribution of free electrons and for microstructure in the distribution that seems to be caused by turbulence in the ISM.

Pulsar dispersion measures, $DM \equiv \int_0^D ds n_e$ (where n_e is the number density of free electrons; D is the pulsar distance), have long been used to deduce the ubiquity of free electrons in the Galaxy. Independent distance information on pulsars, derived from parallax and H I absorption methods and from association with other objects, has shown that line-of-sight average densities are typically $\langle n_e \rangle_{\text{los}} \equiv DM/D \approx 0.01\text{--}0.05 \text{ cm}^{-3}$, the average for objects within a few kiloparsecs of the Sun being 0.025 cm^{-3} (ref. 2). DM/D increases by a factor of two or three for pulsars in the inner Galaxy, as shown by H I absorption-based distance estimates³. Emission measures, $EM \equiv \int_0^{D'} ds n_e^2$, obtained from measurements of H α , radio-recombination lines and thermal radio emission (over path lengths D' generally greater than the pulsar distance D), when combined with DM, yield additional information on the spatial distribution of ionized regions. EM measured towards pulsars⁴, combined with DM, yields volume filling factors $f \gtrsim 0.2$, although these may be biased by the longer path length for EM than for DM ($D' > D$) except, possibly, for pulsars in high-latitude globular clusters⁵.

Radio-wave scattering

Another measurement of electron density can be obtained from radio-wave scattering data^{6–9}, which are sensitive to small-scale fluctuations δn_e on length scales $\sim 10^9\text{--}10^{11} \text{ cm}$. The wave-number spectrum for δn_e consistent with most data is a power law of the form

$$P_{\delta n_e}(k) = C_n^2 k^{-\alpha}, \quad \frac{2\pi}{l_0} \leq k \leq \frac{2\pi}{l_1} \quad (1)$$

where k is the wavenumber, C_n^2 is the spectral coefficient, l_0 and l_1 are the outer and inner scales, and α is empirically determined to be $11/3 \pm 0.3$ for many lines of sight, consistent with the Kolmogorov value of $11/3$ (refs 9–11). The full extent of the spectrum is not known, but recent work^{12–15} suggests

$l_1 \approx 10^7 \text{ cm}$ and $l_0 \gtrsim 10^{14} \text{ cm}$. The lower bound on l_0 follows from refraction phenomena, such as systematic frequency drifts in dynamic spectra, and dispersion-measure changes seen from some pulsars^{13,14,17,18}. Several arguments suggest that l_0 is much larger than these lower bounds, perhaps as large as $\sim 1\text{--}100 \text{ pc}$. These are based on the known structural scales in the ISM⁹, fluctuations in Faraday rotation measures on small angular scales^{19,20}, and the smoothness of the cosmic-ray energy spectrum and its possible relation to a spectrum of magnetic irregularities²¹. Moreover, variations in DM between pulsars in the globular cluster 47 Tucanae (ref. 22) may be fully accounted for by foreground ionized material only if the outer scale is at least 0.1 pc .

We define the scattering measure as $SM = \int_0^D ds C_n^2$, with units $\text{kpc m}^{-20/3}$, a quantity which results directly from radio-wave scattering measurements. For example, an angular broadening diameter of θ_{FWHM} yields $SM = (\theta_{\text{FWHM}}/128 \text{ mas})^{5/3} \nu_{\text{GHz}}^{11/3}$ and $SM = 292(\tau_d/D)^{5/6} \nu_{\text{GHz}}^{11/3}$ for broadening of pulsar pulses by a time τ_d (s) at distance D in kiloparsec (ref. 8). (In these expressions, a medium with uniform C_n^2 is assumed; nonuniform cases must take into account weighting functions that differ for the two kinds of measurements. The value of 128 mas in the first expression applies for extragalactic sources viewed through the ISM; a value of 71 mas should be used for galactic sources. Finally, in very strong scattering, the exponents also change. See ref. 23 for details). We use data for 206 lines of sight, based on angular broadening and scintillation measurements of galactic and extragalactic sources. Figure 1 shows, for pulsars, that SM increases roughly linearly with DM for small DM, but then increases much faster for $DM \gtrsim 100 \text{ pc cm}^{-3}$. As we now show, this variation requires that the distribution of ionized regions be a strong function of galactocentric radius.

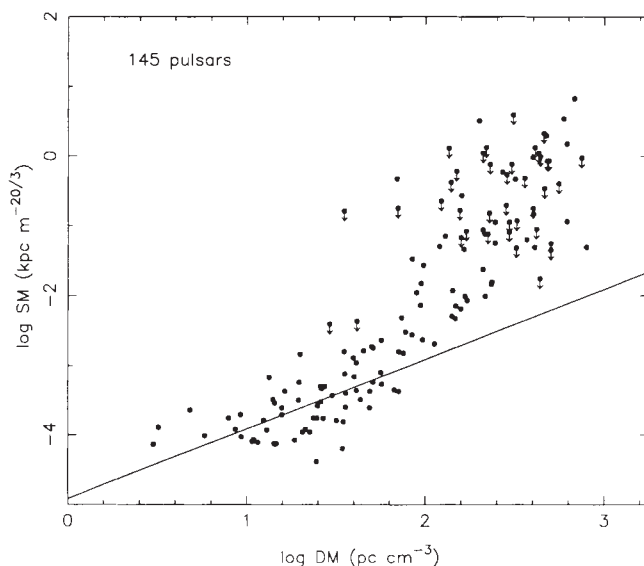


FIG. 1 Scattering measure against dispersion measure, including upper limits on scattering measure. The solid line is a straight-line fit of $\log SM = \log DM + \text{constant}$ to objects with $DM \leq 80 \text{ pc cm}^{-3}$. The data are from references given in ref. 8.

Relationship of DM, EM and SM

Letting the local electron density consist of a mean $\bar{n}_e$ and fluctuation about the mean δn_e , we represent the electron density at location s along the line of sight as $n_e(s) = \bar{n}_e(s) + \delta n_e(s)$ inside ionized regions and $n_e(s) = 0$ outside ionized regions. (Note that we do not assume that ionized regions are necessarily fully ionized.) Thus, we represent fluctuations inside regions using the power spectrum of equation (1), whereas fluctuations from region to region are quantified by the spatial variation of $\bar{n}_e$. The integral of equation (1) over wavenumber space d^3k is simply the mean square of δn_e . It is useful to define the volume-filling factor f of ionized regions and the fractional variation in electron density (locally) as $\varepsilon = \delta n_e(\text{r.m.s.})/\bar{n}_e$. With these definitions, we may write the three measures as

$$\text{DM} = D\langle f\bar{n}_e \rangle \equiv D\langle n_e \rangle_{\text{los}} \quad (2)$$

$$\text{EM} = D\langle (1 + \varepsilon^2)f\bar{n}_e^2 \rangle \equiv \frac{\langle f\bar{n}_e^2 \rangle}{\langle f\bar{n}_e \rangle} \text{DM} + C_{\text{SM}}^{-1} l_0^{\alpha-3} \text{SM} \quad (3)$$

$$\text{SM} = C_{\text{SM}} l_0^{3-\alpha} \frac{\langle f\varepsilon^2 \bar{n}_e^2 \rangle}{\langle f\bar{n}_e \rangle} \text{DM} \equiv \frac{\langle fC_n^2 \rangle}{\langle f\bar{n}_e \rangle} \text{DM} \quad (4)$$

In these expressions, all quantities may vary from region to region (as indicated by the angular brackets which denote averages over the line of sight), except for the spectral index α and the outer scale l_0 , although allowance for variations of the latter would be trivial. We have used the fact that inside ionized regions, the spectral coefficient is $C_n^2 = C_{\text{SM}} \varepsilon^2 \bar{n}_e^2 l_0^{3-\alpha}$ ($C_{\text{SM}} = [(\alpha-3)/2(2\pi)^{4-\alpha}]$) which holds for $\alpha > 3$, and we have assumed the outer scale to be much larger than the inner scale, $l_0 \gg l_1$. The average electron density along the line of sight, $\langle n_e \rangle_{\text{los}} = \langle f\bar{n}_e \rangle \approx 0.025 \text{ cm}^{-3}$, is a useful fiducial value which follows readily from pulsar measurements. We expect that variations in the cloud-to-cloud internal density will follow a distribution such that $\zeta \equiv \langle \bar{n}_e^2 \rangle / \langle \bar{n}_e \rangle^2 \approx 2$. Implicit in the above relations is the assumption that the ionized regions contributing to SM and EM are the same as those that contribute to DM. If this were not the case, the above equations would have to allow for different filling factors for the regions that dominate DM and those that dominate SM. But the available evidence, including that in Fig. 1, is consistent with an interpretation where the dispersing and scattering electrons are one and the same. The strongest evidence is that large-scale variations (over a few parsecs) in δn_e required for fluctuations in Faraday rotation measure^{19,20} and DM fluctuations in globular cluster pulsars²² are consistent with an extrapolation of the wavenumber spectrum from the small scales that cause diffraction effects. This implies, further, that $\delta n_e/\bar{n}_e \approx 1$, requiring that nearly all electrons are consistent with the Kolmogorov spectrum.

Pulsars allow estimation of DM and SM on the same propagation path, making it possible to constrain some of the factors that appear in equations (2)–(4). Equation (4) implies a strict proportionality between SM and DM only if the ratio of averages in the equation is constant, and this evidently is not the case. Objects in Fig. 1 with $\text{DM} \leq 80 \text{ pc cm}^{-3}$ are roughly consistent with simple proportionality (keeping in mind that errors of a factor of two in distance lead to a sizeable error in SM). Objects with larger DM follow a much steeper relation, as might be expected in the inner Galaxy where n_e increases. Increases in n_e alone cannot, however, account for the steep rise of SM for large DM. For the purpose of argument, we will assume that all quantities except $\bar{n}_e$ are independent of location. Then we may write, using $\alpha = 11/3$ here and throughout,

$$\text{SM/DM} = C_{\text{SM}} \langle n_e \rangle_{\text{los}} \left[\frac{\zeta \varepsilon^2}{f l_0^{2/3}} \right] \quad (5)$$

Using $\langle n_e \rangle_{\text{los}} = 0.025 \text{ cm}^{-3}$ and evaluating for small DMs from Fig. 1, we find that $[\zeta \varepsilon^2 / f l_0^{2/3}] \approx 0.2$ (for l_0 in pc). To fit large

TABLE 1 Gaussian plus inner annulus model

Parameter	Fit to SM/DM data†	Fit to distance/DM data‡
$[f\bar{n}_e^{(0)}]_1 (\text{cm}^{-3})$	0.025 ± 0.01	0.025 ± 0.003
$[f\bar{n}_e^{(0)}]_2 (\text{cm}^{-3})$	$0.2^{+0.15}_{-0.10}$	$0.2^{+0.3}_{-0.1}$
$H_1 (\text{kpc})$	$1.0^{+0.5}_{-0.1}$	1.0 ± 0.25
$H_2 (\text{kpc})$	0.15 ± 0.05	0.175 ± 0.13
$A_1 (\text{kpc})$	≥ 20	≥ 20
$A_2 (\text{kpc})$	2 ± 0.25	2 ± 0.5
$\bar{R}_2 (\text{kpc})$	4 ± 0.5	4 ± 1
$\left(\frac{\zeta \varepsilon^2}{f l_0^{2/3}} \right)_1 (\text{pc}^{-2/3})$	0.43 ± 0.1	—
$\left(\frac{\zeta \varepsilon^2}{f l_0^{2/3}} \right)_2 (\text{pc}^{-2/3})$	21.5^{+20}_{-10}	—
$\Sigma_{\text{in}} \S$	2.42	10.8
$\sigma_{\text{out}}/\Sigma_{\text{in}} \parallel$	0.27	0.20

Parameters are defined in equation (6) and related text.

† The fit to SM/DM data constitutes our preferred model for calculating pulsar distances.

‡ The fit to DM and distance data yields no information on the ‘fluctuation’ parameter $\zeta \varepsilon^2 / f l_0^{2/3}$ (see text).

§ Σ_{in} is the r.m.s. value of log SM (that is, $\langle (\log \text{SM})^2 \rangle^{1/2}$) for the SM/DM fit, whereas it is simply the r.m.s. distance (in kpc) for the distance/DM fit.

|| σ_{out} is the r.m.s. difference between model and actual log SM for the second column, and modelled and actual distance in the third.

DM lines of sight toward the inner Galaxy, however, this combination must be 100 times larger if $\langle n_e \rangle_{\text{los}}$ is the same everywhere, or 20 times larger if $\langle n_e \rangle_{\text{los}}$ is five times larger towards the inner Galaxy, an upper bound consistent with the work of ref. 3. We conclude that the rapid increase of SM with DM requires $[\zeta \varepsilon^2 / f l_0^{2/3}]$ to increase systematically as well as randomly in the inner Galaxy compared with values near the solar circle. A comparison²⁴ of SM and EM yields a similar conclusion that scattering increases considerably toward the Cygnus region of the Galaxy.

Galactic distribution

To model the distribution of n_e , we consider two smoothly distributed terms of the form

$$\bar{n}_e(R, z) = \bar{n}_e^{(0)} \exp(-|z|/H_j) \exp[-(R/A_j)^2], \quad j = 1, 2 \quad (6)$$

with corresponding filling factor f_j and fractional variation ε_j , and where R is galactocentric radius and $|z|$ is distance from the galactic plane. Our preferred model has an annular gaussian inner-Galaxy component ($j=2$) which results from replacing R in equation (6) with $R - \bar{R}$, where $\bar{R} = 2A_2$. Each component requires four parameters ($f\bar{n}_e^{(0)}_j$, H_j , A_j , and $(\zeta \varepsilon^2 / f l_0^{2/3})_j$). To fit for the parameters, we performed a grid search using two datasets. First was the set of 67 pulsars for which DM and independent distance data are available²⁶. For these data, we could determine only the first three parameters of each component because DM is insensitive to the fourth. We minimized the mean-square difference between the model distances (calculated from DM) and actual distance constraints. The second data set comprises 206 measurements of SM and either distances (for galactic OH maser sources, Sgr A and so on, and extragalactic sources effectively at infinite distance) or DMs for pulsars (references from ref. 8). For this set, we minimized the mean-square difference between log SM determined from the data and from the model. Remarkably, both datasets yield consistent results, summarized in Table 1, where we show values for the fit parameters. The goodness of fit may be measured by comparing the r.m.s. residual (for log SM and D in the two kinds of fits) with the r.m.s. values before fitting, as given in the table. Details may be found in ref. 8.

The model accounts for the bulk of the variation in SM, although there are large departures from the model. Some of these are due to errors of up to a factor of two in the distance estimates of the pulsars. Most 'discrepant' directions have excess scattering, such as the Vela pulsar, Cyg X-3¹⁶, the Galactic Centre source Sgr A²⁷, and the extragalactic source NGC6334B¹⁵. A few objects show a paucity of scattering, including the pulsars 2016+28, 2020+28 and the millisecond pulsar 1937+21. These objects at $l \approx 55^\circ$ – 70° are probably between the local and Sagittarius spiral arms. We self-consistently excluded objects with strong excess scattering from the final model fit because it was clear that an otherwise good fit was being perturbed by 'special' lines of sight.

The model fitting shows that scattering material lies in two smoothly distributed components. The weak outer Galaxy component with scale height ~ 1.0 kpc and $C_n^2 \approx 10^{-3.5} \text{ m}^{-20/3}$ extends to large galactocentric radii. With existing data, only a lower bound may be put on the $1/e$ radial scale. To determine this scale, low-frequency very-long-baseline interferometry of extragalactic sources is needed at low latitudes ($|b| \leq 1^\circ$). The inner Galaxy component has a scale height ~ 10 times smaller than the outer Galaxy component and has $\sim 10^3$ stronger turbulence. We cannot yet determine whether this inner Galaxy component has an annular or filled distribution. In any case, the turbulence is very strong in the inner Galaxy, and it is probably related to the more intense star formation there. Our model fit places constraints on the filling factors of ionized regions in the three components. In the outer Galaxy component, the line-of-sight average electron density is $f\bar{n}_e \approx 0.025 \text{ cm}^{-3}$ and estimates of $f \geq 0.2$ imply $\epsilon \geq 0.14$ for an outer scale of 1 pc. Estimates of f from a combination of EM and DM may be obtained from $\hat{f} = \zeta(1 + \epsilon^2)[(\text{DM})^2 D^{-1} \text{EM}^{-1}]$, which may be substantially larger than estimates⁴ that assume $\zeta = 1$, $\epsilon = 0$. If ϵ and l_0 are the same everywhere, then the filling factor in the inner Galaxy is at least 50 times smaller than in the outer Galaxy.

In addition to the smooth components, there are 'clumps' of intense turbulence whose scale height is approximately the same as that of the inner Galaxy component. A notable example is the H II complex NG6334, 1.7 kpc from the Sun. The extragalactic source behind this complex shows the largest scattering

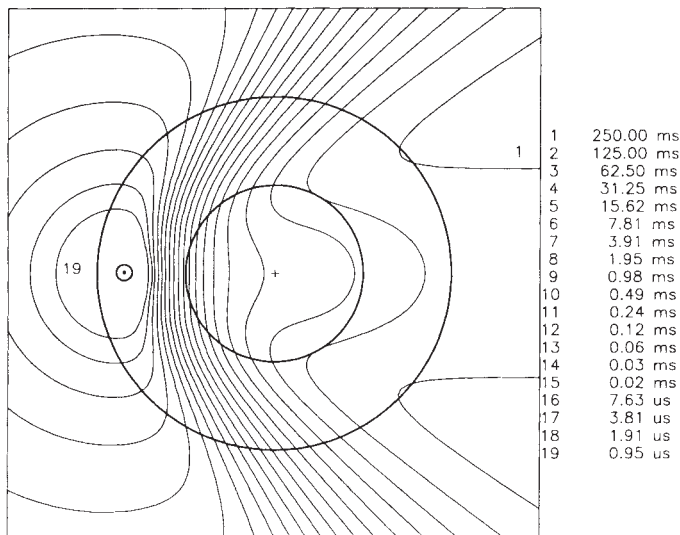


FIG. 2 Contours of temporal broadening time (at 1 GHz) against galactic coordinates. The broadening time scales as $\nu^{-4.4}$ (except for the largest values, for which a ν^{-4} scaling holds²³). The contours are plotted for lines of sight with $|b|=0^\circ$. The large circles are at galactocentric radii of 5 and 10 kpc. The Sun ($\odot$) is at 8.5 kpc. The first and last contours are labelled.

measure of any source¹⁵. Using a size of 10 pc, the internal C_n^2 is $\sim 10^5 \text{ m}^{-20/3}$, many orders of magnitude larger than the inner-Galaxy component. The filling factor f_c of clumps of strong turbulence may be estimated as follows, by assuming these follow a Poisson distribution in the Galaxy with characteristic size R_c . We use the effective distance through a 100-pc scale height for sources with and without excess scattering to estimate the mean free path L_{mfp} for intersecting a clump. For 31 out of 151 sources showing enhanced scattering, we estimate $5 \leq L_{\text{mfp}} \leq 10$ kpc, corresponding to a volume-filling factor for clumps $f_c = 4R_c/3L_{\text{mfp}} \approx 10^{-3.8 \pm 0.2} R_c$ (pc), consistent with the results of ref. 10.

We considered alternative models, such as one where the inner Galaxy component has a maximum at the Galactic Centre rather than having an annular distribution. The available data cannot readily distinguish filled and annular distributions, but we prefer the annular distribution because it may be plausibly related to the molecular ring in the inner Galaxy. We also considered a model where the inner Galaxy component consisted of distinct clumps rather than consisting of a smooth distribution of regions superposed with clumps with much smaller filling factor. Such a model cannot be fitted to the data because the parameters of the strongest clumps seem to be different from those required of more prevalent clumps. In particular, the filling factor for strong clumps f_c is nearly 100 times smaller than the filling factor for the smooth inner-Galaxy component.

Implications

Our model allows better distance estimates for pulsars to be made by integrating n_e until the correct DM is obtained. Pulsar distances toward $|l| \leq 30^\circ$ are generally smaller than those calculated from the model of ref. 28 suggesting that most known pulsars are on 'our' side of the Galactic Centre, to which the dispersion measure is 900 pc cm^{-3} . For those cases where the scattering is higher than predicted by the model, it may be possible to derive improved distance estimates using the scattering measure together with the dispersion measure. Work on improved distance estimates will be reported elsewhere where we will also introduce spiral arm structure into the model. For now, we recommend that equation (6) with parameters in the first column of Table 1 be used to calculate pulsar distances based on dispersion measures.

Temporal broadening of pulsar pulses inhibits detection of pulses from short-period pulsars in the inner Galaxy. Using the inner- and outer-Galaxy components of our model (ignoring clumps) we may derive lower bounds on the temporal broadening. Figure 2 shows contours of broadening at 1 GHz in the galactic plane ($b=0^\circ$). The values may be scaled to other frequencies by the factor $\nu_{\text{GHz}}^{-22/5}$. Figure 2 shows that pulsars with periods shorter than that of the Crab pulsar (33 ms) are undetectable for distances ≥ 6 kpc when looking toward the Galactic Centre. This suggests that searches in the infrared, which can probe to the Galactic Centre without the deleterious effects of pulse broadening may be the best way to search for young, rotation-driven pulsars.

Figure 3 shows the scattering diameter at 1 GHz predicted by our model. Broadening at other frequencies follows a $\nu_{\text{GHz}}^{-11/5}$ scaling. (Again, the angular broadening is a lower bound because it does not include clumps). Our results are consistent with recent studies²⁹ of interplanetary scintillations at 0.327 GHz, which require ~ 1 arcsec of scattering toward the inner Galaxy ($|l|, |b| \leq 5^\circ$) to explain the absence of scintillating sources in these directions. Our model does not account for the angular broadening of the Galactic Centre source Sgr A* (ref. 27). It is not known if the enhanced scattering is caused by material within the Galactic Centre itself or somewhere in the foreground. If similar or higher scattering is found towards other sources far from the direction of the Galactic Centre, this would suggest a foreground interpretation.

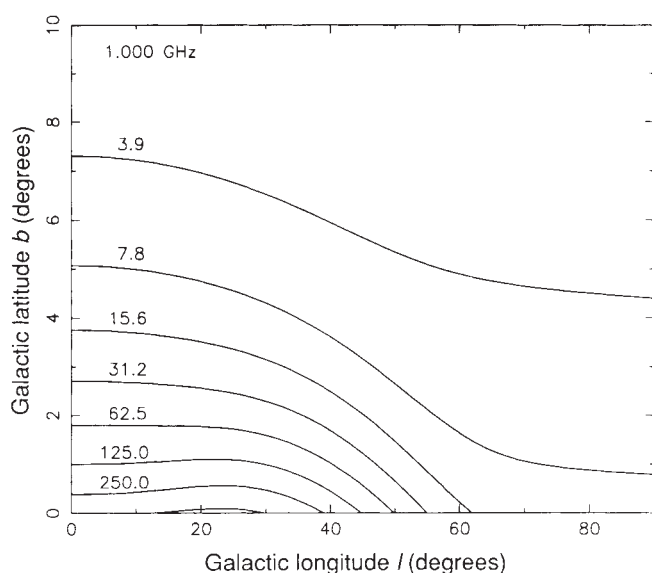


FIG. 3 Contours of angular broadening (full width at half maximum at 1 GHz) of extragalactic sources, against galactic longitude and latitude. Broadening scales as $\nu^{-2.2}$ (but for the heaviest scattering, it will scale as ν^{-2} if the inner scale $l_1 \approx 100$ km (refs 12, 15, 16, 23)). The contours (labelled in milliarcsec) are based on the model described in the text and exclude strong clumps of turbulence which enhance the scattering along some lines of sight. The angular broadening values here are therefore lower bounds for most directions, although gaps in the scattering medium suggest that the contours will overestimate the scattering for other directions.

The scattering measure implies lower bounds (from the second term of equation (3)) on the emission measure and optical depth due to free-free absorption:

$$\text{EM}_{\text{SM}} \approx 544 \text{ pc cm}^{-6} \left(\frac{\text{SM}}{\text{kpc m}^{-20/3}} \right) \left(\frac{l_0}{1 \text{ pc}} \right)^{2/3} \quad (7)$$

The actual EM will be at least twice this value (because $\delta n_e \leq \bar{n}_e$), yielding an optical depth $\tau_{\text{ff}} \approx 10^{-3.3} \text{SM} \nu^{-2.1}$ for ν in GHz, SM in usual units, and for a temperature of 10^4 K. Our model for SM implies that the Galactic centre will be opaque at a frequency no lower than 80 MHz, consistent with the appearance of the Sgr A region in absorption at 80 MHz (ref. 30).

It is commonly assumed that signals from extraterrestrial intelligence would be in the form of slowly modulated, narrow-band signals from transmitters of, perhaps, planetary size or smaller. Such signals will undergo intensity scintillations identical to those seen from pulsars but with characteristic timescales (of the order of hours) longer than those from pulsars because the space velocities of pulsars are an order of magnitude larger than those expected for planetary orbits around main-sequence stars of solar type. Scintillations comprise 100% variations for sources in the strong scattering regime which, from our model, will hold for $\nu \leq 5 \text{ GHz } D_{\text{kpc}}^{11/17}$. Additional implications are discussed in ref. 23.

ISM structure and ionization

Our model has implications for the structure and ionization of the ISM, and for the distribution of interstellar turbulence and the diffusion of cosmic rays. First, the outer Galaxy component

extends to well beyond the solar circle, suggesting that sources of both ionization and turbulence in this component also extend to the outer Galaxy. The increase in density modulation toward the inner Galaxy is substantially larger than the increase in gas density. Therefore, one must look for mechanisms that are nonlinear in the gas density. It is likely that regions of intense δn_e and small filling factor are associated with interfaces of shocks and gas clouds. Supernova and stellar-wind shocks occur at a rate proportional to the local star formation rate $\dot{n}_* \propto n_{\text{gas}}^m$ with $1 \leq m \leq 2$, or possibly with even larger m if regions are near a star formation threshold³¹. Ionized fronts occur at a rate proportional to n_{gas}^{m+1} , so a factor of 5 increase in gas density may account for the increased variation of electron density in the inner Galaxy. Such an interpretation may account for the properties of the two smoothly distributed components we have modelled. Clumps of enhanced turbulence then correspond to regions with especially energetic or salient activity, perhaps on the tail of a distribution of regions that account for the statistical properties of the smooth components.

Electron-density variations are probably related to or driven by magnetic turbulence¹⁹⁻²¹ with δn_e playing the part of a passive tracer of turbulent energy^{32,33}. The spatial distribution and spectrum of density irregularities therefore trace the same quantities for magnetic irregularities. The strong dependence of these quantities on galactocentric radius and distance from the galactic plane, combined with the occurrence of clumps of extremely strong turbulence, should be useful input for studies of cosmic ray propagation in the Galaxy and for observations of γ -ray emission from cosmic-ray interactions with clouds. $\square$

Received 8 July; accepted 15 October 1991.

- Kulkarni, S. R. & Heiles, C. in *Galactic and Extragalactic Radio Astronomy* (ed. Verschuur, G. L. & Kellermann, K. I.) 95-153 (Springer, New York, 1988).
- Weisberg, J. M., Rankin, J. M. & Boriakoff, V. *Astr. Astrophys.* **186**, 307-311 (1987).
- Frail, D., Cordes, J. M., Weisberg, J. M. & Hankins, T. H. *Astrophys. J.* (submitted).
- Reynolds, R. J. *Astrophys. J.* **216**, 433-439 (1977).
- Reynolds, R. J. *Astrophys. J.* **372**, L17-20 (1991).
- Cordes, J. M., Rickett, B. J. & Backer, D. C. (eds) *Radio Wave Scattering in the Interstellar Medium*. *AIP Conf. Proc.* 174 (American Institute of Physics, New York, 1988).
- Rickett, B. J. *Ann. Rev. Astr. Astrophys.* **28**, 561-605 (1990).
- Cordes, J. M., Spangler, S. R., Weisberg, J. M. & Ryan, M. *Astrophys. J.* (submitted).
- Armstrong, J. W., Cordes, J. M. & Rickett, B. J. *Nature* **291**, 561-564 (1981).
- Cordes, J. M., Weisberg, J. M. & Boriakoff, V. *Astrophys. J.* **288**, 221-247 (1985).
- Lee, L. C. & Jokipii, J. R. *Astrophys. J.* **206**, 735-743 (1976).
- Spangler, S. R. & Gwinn, C. R. *Astrophys. J.* **353**, L29-32 (1990).
- Cordes, J. M., Wolszczan, A., Dewey, R. J., Blaskiewicz, M. & Stinebring, D. *Astrophys. J.* **349**, 245-261 (1990).
- Phillips, J. A. & Wolszczan, A. *Astrophys. J.* (submitted).
- Moran, J. M., Rodriguez, L. R., Greene, B. & Backer, D. C. *Astrophys. J.* **348**, 147-152 (1989).
- Molnar, L. et al. *Astrophys. J.* (submitted).

- Smith, F. G. & Wright, W. C. *Mon. Not. R. astr. Soc.* **214**, 97-107 (1985).
- Hewish, A., Wolszczan, A. & Graham, D. A. *Mon. Not. R. astr. Soc.* **213**, 167-179 (1985).
- Simonetti, J. & Cordes, J. M. *Astrophys. J.* **216**, 134-138 (1977).
- Lazio, J., Spangler, S. R. & Cordes, J. M. *Astrophys. J.* **363**, 515-527 (1990).
- Jokipii, J. R. *Astrophys. J.* **372**, 48-57 (1991).
- Manchester, R. N. et al. *Nature* **352**, 219-222 (1991).
- Cordes, J. M. & Lazio, T. J. *Astrophys. J.* **376**, 123-134 (1991).
- Spangler, S. R. & Reynolds, R. J. *Astrophys. J.* **361**, 116-120 (1990).
- Fey, A. L., Spangler, S. R. & Cordes, J. M. *Astrophys. J.* **372**, 132-160 (1991).
- Frail, D. A. & Weisberg, J. M. *Astrophys. J.* **100**, 743-757 (1990).
- Backer, D. C. *Astrophys. J.* **372** (1991).
- Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613-623 (1985).
- Rao, A. P. & Slee, O. B. *Mon. Not. R. astr. Soc.* **234**, 853-862 (1988).
- Kassim, N. E., Erickson, W. C. & LaRosa, T. N. in *The Galactic Center*, *AIP Conf. Proc.* 155, 196-201 (American Institute of Physics, New York, 1987).
- Kennicutt, R. C. *Astrophys. J.* **344**, 685-703 (1989).
- Higdon, J. C. *Astrophys. J.* **309**, 342-364 (1986).
- Montgomery, D., Brown, M. R. & Matthes, W. H. *J. geophys. Res.* **92**, 282-284 (1987).

ACKNOWLEDGEMENTS. This work was supported by the NSF, NASA and the National Astronomy and Ionosphere Center, which operates Arecibo Observatory under contract with the NSF.

A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*

Herman P. Spaink, Douglas M. Sheeley*, Anton A. N. van Brussel, John Glushka†, William S. York†, Teun Tak, Otto Geiger, Eugene P. Kennedy‡, Vernon N. Reinhold* & Ben J. J. Lugtenberg

Leiden University, Department of Plant Molecular Biology, Nonnensteeg 3, 2311 VJ Leiden, The Netherlands

* Harvard School of Public Health, Department of Nutrition, Boston, Massachusetts 02115, USA

† Complex Carbohydrate Research Center, 220 Riverbend Road, Athens, Georgia 30602, USA

‡ Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, Boston, Massachusetts 02115, USA

In *Rhizobium leguminosarum* biovar *viciae*, the *nodABC* and *nodFEL* operons are involved in the production of lipo-oligosaccharide signals which mediate host specificity. The structure of these metabolites and those produced in *nod* mutants links the *nodE* and *nodL* genes to specific chemical features of the signal molecules. A *nodE*-determined, highly unsaturated fatty acid and a *nodL*-determined *O*-acetyl substituent are essential for the ability of the signals to induce nodule meristems on the host plant *Vicia sativa*.

Sym plasmid (pRL1J1), Sym plasmids with *nod* gene mutations, or cloned *nod* genes (Fig. 1). The HPLC elution patterns using ultraviolet light detection (206 nm) are correlated with the radioactively labelled Nod metabolites detected after TLC. For example, in the situation of the wild-type Sym plasmid the three major radioactive spots seen after TLC (indicated by lines in Fig. 1a) were separated by HPLC (Fig. 1b) in four peaks each preceded by a satellite peak. The satellite peaks are the result of the isomeric separation of the α - and β -anomeric forms of the lipo-oligosaccharides (Fig. 2).

General characterization

The Nod metabolites, purified to homogeneity, were further analysed by mass spectrometry using fast-atom bombardment ionization (FAB) (Fig. 2 and see also Fig. 1). From the results of these experiments we conclude that: (1) in wild-type *Rhizobium* from the *nodABC* operon, at least *nodC* is required for the production of four Nod metabolites with (nominal) relative molecular masses (M_r) of 1,291, 1,088, 1,297 and 1,094; (2) the presence of the NodE protein is required for the production of the two Nod metabolites with M_r of 1,291 and 1,088; (3) the two other flavonoid-inducible metabolites with M_r of 1,297 and 1,094 are still produced in the absence of NodE; (4) in the absence of the NodL protein four Nod metabolites were purified with M_r of 1,249, 1,046, 1,255 and 1,052, which indicates the absence of an acetyl group compared with the wild-type Nod metabolites; and (5) when only *nodABC* and the regulatory *nodD* genes are present, two Nod metabolites were purified with M_r of 1,255 and 1,052.

The FAB spectra (Fig. 2) and additional chemical analyses of the isolated compounds indicated all to be lipo-oligosaccharides similar to the signal molecule NodRm-1 of *R. meliloti* discovered by Lerouge *et al.*⁸. But several structural differences of the *R. leguminosarum* bv. *viciae* Nod metabolites with the NodRm-1 molecule were observed: (1) the Nod metabolites of *R. leguminosarum* are devoid of sulphate; (2) they contain an additional *O*-acetyl moiety; (3) their *N*-acyl group is different and (4) the number of *N*-acetylglucosamine units of the sugar backbone of some of the Nod metabolites is different. Furthermore, the two NodFE-related Nod metabolites had a characteristic ultraviolet spectrum (as demonstrated by the HPLC diode array analysis) with an absorption maximum at 303 nm.

Chemical structures

The structure of several Nod metabolites was further analysed. Nod metabolites were purified from 200 l of rhizobial culture. To obtain a higher yield of Nod metabolites, *R. leguminosarum* strains in which the appropriate *nod* genes were introduced in a high copy number were used, resulting in about threefold overproduction of the Nod metabolites. From the detailed chemical (Fig. 2) and NMR analyses (Fig. 3) of the isolated Nod metabolites from these strains we determined their

THE host-specific interaction between *Rhizobium* bacteria and leguminous plants leads to nitrogen-fixing root nodules. Various rhizobial genes have been identified which determine the host specificity of nodulation¹. The *nodE* gene in the *nodFEL* operon is the main factor in determining the difference in host-specific characteristics of the *R. leguminosarum* biovars *viciae* and *trifolii*². The NodE protein is homologous to a family of β -ketoacyl synthases including the FabB protein of *Escherichia coli*³, whereas NodF is homologous to acyl carrier proteins⁴. The NodF protein contains a 4'-phosphopantetheine prosthetic group, indicating its role in the production of β -ketide-derived compounds⁵. The NodL protein is homologous to LacA of *E. coli*⁶, which has the enzymatic activity of a transacetylase⁷.

The common rhizobial genes *nodABC* must be present for the production of acylated glucosamine oligosaccharide signals⁸⁻¹². Here we report that, besides *nodABC*, no other Sym plasmid localized genes are required for the production of such lipo-oligosaccharide molecules. We also show that the NodE protein of *R. leguminosarum* biovar *viciae* is involved in the production of a novel, highly unsaturated lipid moiety on the oligosaccharide molecules. This novel lipid, and an additional *O*-acetyl modification produced only in the presence of the NodL protein, seem to be essential for the lipo-oligosaccharides to be able to induce the formation of nodule meristems and the production of new flavonoids¹⁵⁻¹⁷ *in vitro* on the host plant *Vicia sativa*.

Metabolite purification

Using radioactive acetate as a precursor, it was shown that the regulatory *nodD* gene as well as the *nodABC* and *nodFEL* operons of *R. leguminosarum* bv. *viciae* are involved in the production of at least five metabolites which are excreted into the growth medium^{11,13}. Thin layer chromatography (TLC) and high-performance liquid chromatography (HPLC) were used as a test system for the purification of the excreted radiolabelled Nod metabolites from a *Rhizobium* strain containing a wild-type

TABLE 1 Biological activity of Nod metabolites

Producing strain	Tested compound	HAD	TSR	INI	Nodule meristems
RBL5560 (wild-type Sym plasmid)	NodRlv-V(Ac, C18:4)	+10 ⁻¹¹	+10 ⁻¹⁰	+5 × 10 ⁻⁸	+5 × 10 ⁻⁸
	NodRlv-IV(Ac, C18:4)	+10 ⁻¹¹	+10 ⁻¹⁰	+5 × 10 ⁻⁸	+5 × 10 ⁻⁸
	NodRlv-V(Ac, C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	—	—
	NodRlv-IV(Ac, C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	—	—
RBL601 (<i>nodE::Tn5</i>)	NodRlv-V(Ac, C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	—	—
RBL5793 (<i>nodL::Tn5phoA</i>)	NodRlv-IV(Ac, C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	—	—
	NodRlv-V(C18:4)	+10 ⁻¹¹	+10 ⁻¹⁰	—	—
	NodRlv-IV(C18:4)	+10 ⁻¹¹	+10 ⁻¹⁰	—	not tested
	NodRlv-V(C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	not tested	not tested
LPR5045.pMP292 (cloned <i>nodDABC</i>)	NodRlv-IV(C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	not tested	not tested
	NodRlv-V(C18:1)	+(10 ⁻¹¹)	+(10 ⁻¹⁰)	—	not tested
RBL5560 (wild-type Sym plasmid)	Crude preparation of Nod metabolites*	+	+	+	+
ANU843 (wild-type <i>R. leguminosarum</i> bv. <i>trifolii</i>)	Crude preparation of Nod metabolites*	+	+	—	—
	<i>N, N', N'', N'''</i> -tetraacetylchitotetraose	—	—	—	—

Nod metabolites and the chitin fragment were tested in dilution series on *Vicia sativa* subsp. *nigra* for the indicated phenotypes in the presence and absence of the mild detergent 3-((3-cholamidopropyl)dimethyl-ammonio)-1-propanesulphonate (CHAPS, Sigma). Plus and minus indicate whether a significant positive reaction was observed using the highest tested concentration of 5×10^{-7} M. Also indicated is the minimal concentration (in M) at which the phenotype was still clearly observed (see also Fig. 5). Values indicated between parentheses indicate that the effect was only observed at that concentration in the presence of 0.01% CHAPS. In all cases the Nod metabolites induced a similar HAD and TSR response in the presence of CHAPS at 10^{-11} and 10^{-10} M concentrations, respectively. In the absence of CHAPS the Nod metabolites with the C18:1 substituent induced only a significant HAD response at 5×10^{-8} M. Tests for root-hair deformation (HAD), thick and short roots (TSR), and increased *nod* gene inducing activity (INI) were done as described previously^{14,15}. HAD and TSR were scored after 7 days of incubation and INI was scored after 4 days of incubation using RBL5284 as a test strain¹⁵.

* Nod metabolites produced by *R. leguminosarum* bv. *viciae* strain RBL5560 and *R. leguminosarum* bv. *trifolii* strain ANU843 (ref. 2) after induction with naringenin were isolated from 0.1 l cultures grown in B⁻ medium² by *n*-butanol extraction of the spent culture medium. The presence of the major Nod metabolites in the *n*-butanol extract of ANU843 was determined by TLC analysis of produced radiolabelled Nod metabolites¹³. The relative quantity (compared with strain RBL5560) of Nod metabolites produced by strain ANU843 was estimated on the basis of the radioactive labelling studies. An INI phenotype or nodule meristem induction was not observed with the Nod metabolites of ANU843, not even in concentrations estimated to be eight-fold higher than the concentration of Nod metabolites in the used crude preparation of strain RBL5560. The same result was obtained with strain RBL601 (Fig. 1) harbouring plasmid pMP263 (which contains the *nodE* gene of *R. leguminosarum* biovar *trifolii*)².

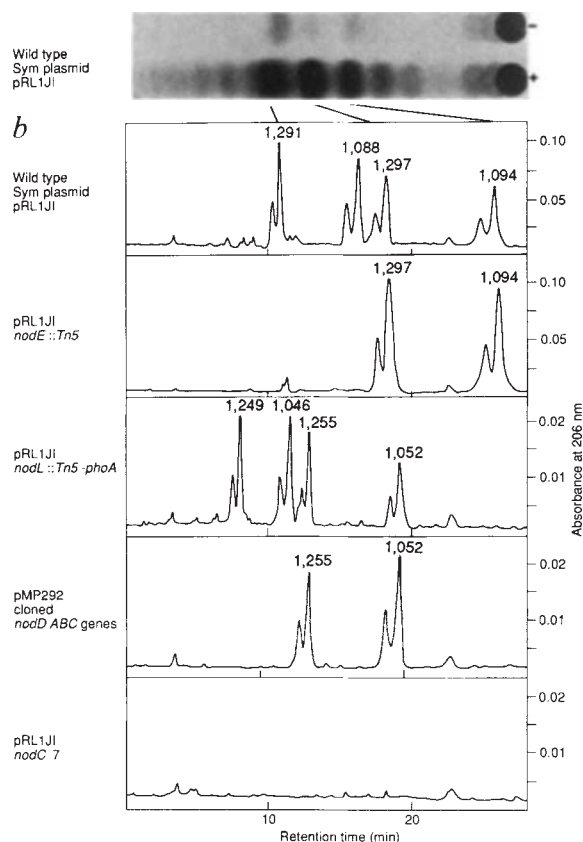


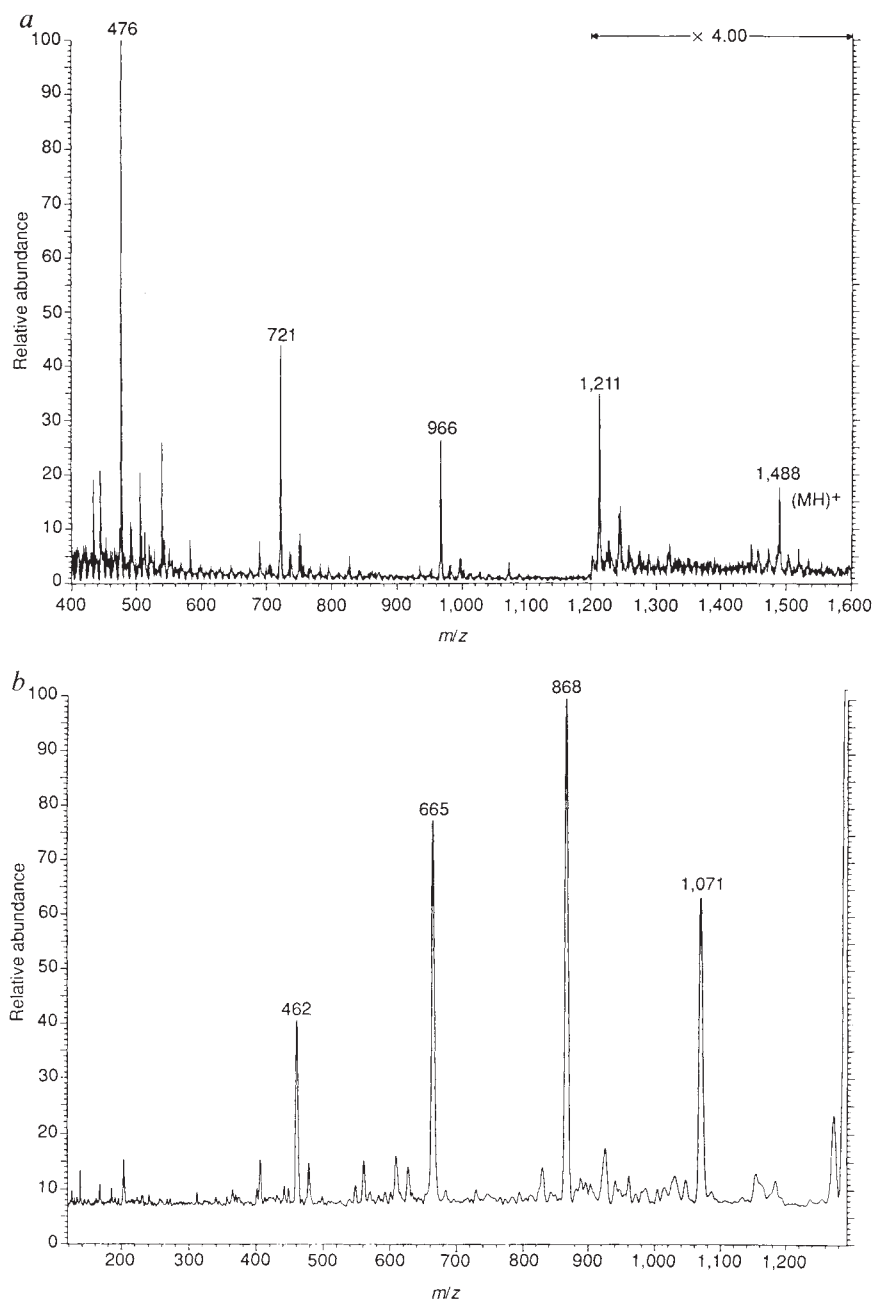
FIG. 1 Detection and purification of *R. leguminosarum* bv. *viciae* Nod metabolites. The *Rhizobium* strains used were derived from the Sym plasmid-cured strain LPR5045 (ref. 26) by the introduction of the wild-type Sym plasmid pRL1Jl, *nod* mutants from this Sym plasmid, or cloned *nod* genes. The resulting strains were: RBL5560 (wild-type Sym plasmid pRL1Jl *mep*:

Tn5)²⁷, RBL601 (pRL1Jl *nodE::Tn5*)²⁸, RBL5793 (pRL1Jl *nodL::Tn5phoA*)¹⁵, LPR5045.pMP292 (cloned *nodDABC* genes of pRL1Jl as a *Clal*-*EcoRI* fragment in pMP92)²⁷, and RBL607 (pRL1Jl::Tn5 *nodC7*; a nonpolar deletion in the *nodC* gene)²⁸. Cells of these strains were grown in 20 ml B⁻ medium² in the presence of ¹⁴C-labelled acetate (0.5 mCi, specific activity 50 mCi mol⁻¹) and the inducer naringenin (1 μ M) as described¹³. As a control strain RBL5560 was grown under identical conditions in the absence of naringenin. *n*-butanol extracts containing the radiolabelled Nod metabolites were prepared from the spent growth medium¹³. *a*, Separation and detection of radiolabelled Nod metabolites using thin layer chromatography (TLC). Aliquots (1/500) of the *n*-butanol samples derived from the uninduced (–) and induced (+) strain RBL5560 were applied to octadecyl silica TLC plates (Sigma) which were developed in 50/50 acetonitrile/water. *b*, Separation and detection of Nod metabolites from various *R. leguminosarum* strains on HPLC. The radioactive *n*-butanol extracts were mixed with *n*-butanol extracts from the spent nonradioactive growth medium of 5 l of culture from the same strain grown under similar conditions. Using the TLC system described above as a test system, the flavonoid-inducible radioactive metabolites of these strains were purified from the extracts using liquid chromatography on silica gel 60 (using a 82/18 to 60/40 acetonitrile/water block gradient), Sephadex-LH20, and ion exchange resin. In these chromatography systems, the major flavonoid-inducible Nod metabolites (indicated with lines in *a*) chromatographed as a single peak. As a final purification step, 1/20 of this material was applied to a reversed phase HPLC column (5 μ m, 4 × 250 mm, Pharmacia LKB) with 30/70 acetonitrile/water as a mobile phase (flow rate, 0.7 ml min⁻¹). The metabolites were fractionated into compounds of increasing hydrophobicity by elution with 40/60 acetonitrile/water. The ultraviolet absorbance profile (206 nm) of the eluate is shown in *b*. Fractions (0.3 ml) were collected and aliquots were analysed for radioactivity and tested by the described TLC system (*a*). The results showed that the radioactivity was correlated with the ultraviolet absorbance and that the detected peaks could be related to the radioactive flavonoid-inducible spots detected in TLC¹³. As an example, the relation between TLC and HPLC data obtained with the wild-type is indicated (*a*) showing that the metabolites with *M_r* 1,088 and 1,297 comigrated as one single spot in TLC (*a*). The obtained HPLC fractions of interest were subsequently analysed by chemical degradation and FAB-mass spectrometry (Fig. 2). The expected nominal *M_r* deduced from these results are indicated above the corresponding HPLC peaks. Purification methods are described in detail in the supplementary information.

complete primary structure (Fig. 4). The most remarkable structural feature of these samples was the lipid moiety of the NodFE-related products. This seems to be a highly unsaturated C18 fatty acid. Three *trans* double bonds are in conjugation with the carbonyl group and probably account for the observed 303 nm absorption maximum. A fourth *cis* double bond is located at position 11. This novel fatty acid was not detected in any of the other studied Nod metabolites of *R. meliloti*⁸, *R. leguminosarum* bv. *trifolii* (A. Aarts, personal communication) or *Bradyrhizobium japonicum* (H.P.S., R. Carlson and G. Stacey, unpublished results). In agreement with the proposed role of the NodF and NodE proteins in the production of β -ketides^{3-5,11}, these proteins are probably involved in the synthesis of the highly unsaturated fatty-acid moiety. In the absence of NodE, only a monounsaturated C18 fatty-acid moiety was detected (Fig. 4). This fatty acid (*cis*-vaccenic acid), which contains a *cis* double bond at position 11, was also the most

abundant fatty acid present in cells of the *R. leguminosarum* strain LPR5045 (data not shown) which is the chromosomal background used throughout this research (Fig. 1). Another feature of the *R. leguminosarum* bv. *viciae* Nod metabolites is the presence of an *O*-acetyl group on the 6 position of the nonreducing terminus. As this group was absent in a strain lacking the *nodL* gene, we assume NodL to be involved in its addition. This is consistent with the observed homology of the NodL protein with the LacA protein of *E. coli*⁶ which has the enzymic activity of a transacetylase⁷. Thus the results show that of the *R. leguminosarum* *nod* genes only the *nodABC* genes are required to produce a basic lipo-oligosaccharide structure. The *nodFEL* genes are required to produce Nod metabolites with a different *N*-acyl chain and an additional *O*-acetyl substituent. The nomenclature used (Fig. 4) is similar to that now being used for the *R. meliloti* signal compounds (J. Dénarié and J. C. Promé, personal communication).

FIG. 2 Mass spectrometry and chemical analysis of Nod metabolites. The HPLC fractions containing the Nod metabolites (Fig. 1) were analysed by gas chromatography-mass spectrometry (GC-MS), after acid hydrolysis. This indicated 2-deoxy-2-amino glucose to be the only saccharide present. Direct FAB-MS analyses provided the following information: for the major HPLC peaks, pseudomolecular ions ($M+H$)⁺ were recorded at $m/z=1,293, 1,090, 1,299$ and $1,096$ for the wild-type Nod metabolites (inferred $M_r=1,291, 1,088, 1,297$ and $1,094$); at $m/z=1,299$ and $1,096$ for the *nodE* mutant (inferred $M_r=1,297$ and $1,094$); at $m/z=1,251, 1,048, 1,257$ and $1,054$ for the *nodL* mutant (inferred $M_r=1,249, 1,046, 1,255$ and $1,052$) and at $m/z=1,257$ and $1,054$ for the strain containing the cloned *nodABC* genes (inferred $M_r=1,255$ and $1,052$), in order of increasing hydrophobicity, respectively. The satellite peak preceding each major HPLC fraction provided an identical FAB spectrum, suggesting an isomeric separation of the α - and β -anomers, expected of reducing oligosaccharides under these conditions. This was also confirmed by the NMR analyses. The samples were also analysed as methyl, acetyl and NaBH₄-reduced derivatives using methods described previously²⁹⁻³². An example is given for the permethylated derivative of the wild-type Nod metabolite with $M_r=1,291$ (a). For all analysed compounds, the relationships between the parent ions and their fragments indicate an *N*-acetylated amino sugar homopolymer composed of four or five residues, with a variable acyl modification on the nonreducing terminus only. The data also indicate the presence of an additional *O*-acetyl group at the nonreducing end sugar of the Nod metabolites produced by the strains containing the *nodL* gene. The sequential parent ion relationship was corroborated by ($M+H$)⁺ focusing and fragmentation by collision-induced dissociation (CID). An example is shown for the ion at $m/z=1,293$ (wild-type strain) using the B/E-linked scan method (b). Linkage analysis was done on the $M_r=1,291$ and $M_r=1,297$ metabolites by the production and analysis of partially methylated alditol acetates as described previously³². The results indicated 1,4-linkages between all monosaccharide moieties. The Nod metabolites with $M_r=1,297$ and $M_r=1,094$ (isolated from the wild-type and the *nodE* mutant) released 11-octadecenoic acid and a small quantity of octadecanoic acid after complete hydrolysis (aqueous 5 M HCl, 100 °C, 5 h) as judged by the GC-MS analysis of their methyl esters. Mild methanolysis (1 M HCl in water-free methanol, 80 °C, 1 h) followed by trimethylsilylation, resulted in the production of the TMS methylglycosides of C18:1-*N*-acylglucosamine as indicated by the appearance of characteristic fragments in the GC-MS analysis¹⁰. **METHODS.** For GC, use was made of a DB-1 capillary column (30 or 60 m,



J & W Scientific). Mass spectra were recorded on a VG ZAB-SE high-performance mass spectrometer (VG Analytical, Manchester) as described previously²⁹. More details are described in the supplementary information.

Biological activity

Nod metabolites excreted by *R. leguminosarum* bv. *viciae* elicit various responses on the roots of their host plants^{11,14,15}. Purified Nod metabolites (Fig. 1) were added to the roots of *V. sativa* subsp. *nigra* plants in a series of concentrations which were subsequently analysed for root-hair deformation (HAD)¹⁴, thick and short roots (TSR)¹⁴, and the production of new flavonoids (INI)¹⁵⁻¹⁷. As a control a tetramer of β -1,4-linked *N*-acetylglucosamine (*N*, *N'*, *N''*, *N'''*-tetraacetylchitotetraose)¹⁸ was tested. This small chitin fragment is identical with the sugar backbone of two of the wild-type Nod metabolites without the *N*-acyl and *O*-acetyl substituents. To circumvent the risk of insolubility of the Nod metabolites in water we also tested their effect and that of the control in the presence of a mild detergent. The results (Table 1) indicate all tested Nod metabolites to be active in less than nanomolar concentrations in the HAD and TSR assay. By contrast, the INI phenotype required the presence of the NodFE-dependent lipid moiety as well as the presence of the *O*-acetyl substituent. To test whether important signal molecules were lost in any of the purification steps we also tested all fractions after the chromatographic separation on silica gel and reversed-phase HPLC (Fig. 1) for induction of HAD, TSR and INI on *V. sativa*. The results showed that the mentioned biological effects induced by extracts from the spent culture medium are completely correlated with the presence of Nod metabolites as measured by radioactivity and ultraviolet spectroscopy (data not shown).

Cytological studies of 'empty' nodules induced by *R. meliloti* mutants which do not invade efficiently, suggested that wild-type bacteria are able to elicit, at a distance, cortical cell dedifferentiation and initiation of nodule organogenesis by means of a 'nodule-inducing principle'^{19,20}. These observations led us to test whether the purified extracellular Nod metabolites of *R. leguminosarum* bv. *viciae* (Fig. 1) can elicit nodule meristem

formation *in vitro* using the staining method developed by Truchet *et al.*²¹. Intensive microscopic studies, in which care was taken to distinguish nodule meristems from starting lateral root meristems, revealed that the purified NodFE-related wild-type signal molecules NodRlv-V(Ac, C18:4) and NodRlv-IV(Ac, C18:4) (Fig. 4) can indeed induce cortical cell dedifferentiation on *V. sativa* indistinguishable from the first stage of normal nodule organogenesis (Fig. 5). By contrast, none of the other tested Nod metabolites could elicit nodule meristem induction (Table 1). A mixture of Nod metabolites produced by *R. leguminosarum* bv. *trifolii* strain ANU843 was unable to produce nodule meristems on *V. sativa* even in concentrations estimated to be eightfold higher than the active concentrations for the *R. leguminosarum* bv. *viciae* signal molecules (Table 1).

Discussion

We have shown that for eliciting nodule meristems both the novel fatty-acid moiety (*nodE*-dependent) and the *O*-acetyl substituent (*nodL*-dependent) are crucial for the function of the signal molecules. In the *in vitro* test system, as *in vivo*, the *nodE* gene of *R. leguminosarum* bv. *trifolii* is not able to complement for the function of the bv. *viciae* *nodE* gene, also not in an isogenic background (Table 1). The involvement of Node in the production of the novel lipid is therefore apparently the major step in distinguishing the host range of the *R. leguminosarum* bv. *viciae* and *trifolii*². By contrast, in *R. meliloti* the presence of a sulphate group on the reducing-end sugar of NodRm-1 is a major determinant of host specificity⁹. This sulphate substituent is produced by means of the *nodH* and *nodPQ* gene products^{9,22}. A difference in the chain length of the oligomer backbone of the signal molecules as in *R. leguminosarum* bv. *viciae* (Fig. 4) was also discovered in *R. meliloti* (J. Dénarié, personal communication; S. Long, personal communication). But the biological importance of the variability

FIG. 3 NMR spectrometry of Nod metabolites. *a*, Regions showing acyl proton signals from a 600 MHz DQF-COSY spectrum of the wild-type Nod metabolite with $M_r=1,291$ (NodRlv-V(Ac, C18:4)) in DMSO. The six signals between δ 5.8 and δ 7.0 correspond to methine protons from positions 2 to 7 of the acyl chain, and show connectivity to the upfield signals of the methylene protons. The large 3J proton-proton coupling constants (16 Hz) between H2 and H3, and H6 and H7, indicate the *E* configuration, and a value of 13.2 Hz between H4 and H5 also suggests the same. The carbon chemical shift (supplementary information S5) of δ 31.8 for C8 confirms the *trans* double bond between C6 and C7 (ref. 33). The connectivity from H8 to methine signals at δ 5.3, as well as the integration of the methylene proton signals places the fourth double bond between C11 and C12. The *Z* configuration is based on the carbon chemical shifts of δ 26.2 and δ 26.4 for the adjacent methylenes (ref. 33). The equivalent spectra for the Nod metabolite with $M_r=1,297$ (NodRlv-V(Ac, C18:1), supplementary information S7 and S8) show a single set of methine signals at δ 5.3 attributable to one *cis* double bond. The *Z* configuration of this bond is based on the carbon chemical shifts δ 27.1 for the adjacent methylene groups. The exact position on the chain cannot be determined from the NMR data, because of overlapping methylene signals. *b*, The sugar region from a DQF-COSY spectrum of the Nod metabolite with $M_r=1,297$ (NodRlv-V(Ac, C18:1)) in DMSO. Two downfield signals at δ 4.26 and δ 4.03 were assigned to H6 and H6' of the 6-*O*-acetyl substituted glucosamine¹⁰, as were H5 and H4, at δ 3.41 and δ 3.09, respectively. From the HMQC spectrum (supplementary information S6 and S9), the upfield chemical shift position for the corresponding C4 (δ 71.0) confirmed the residue to be at the nonreducing terminus.

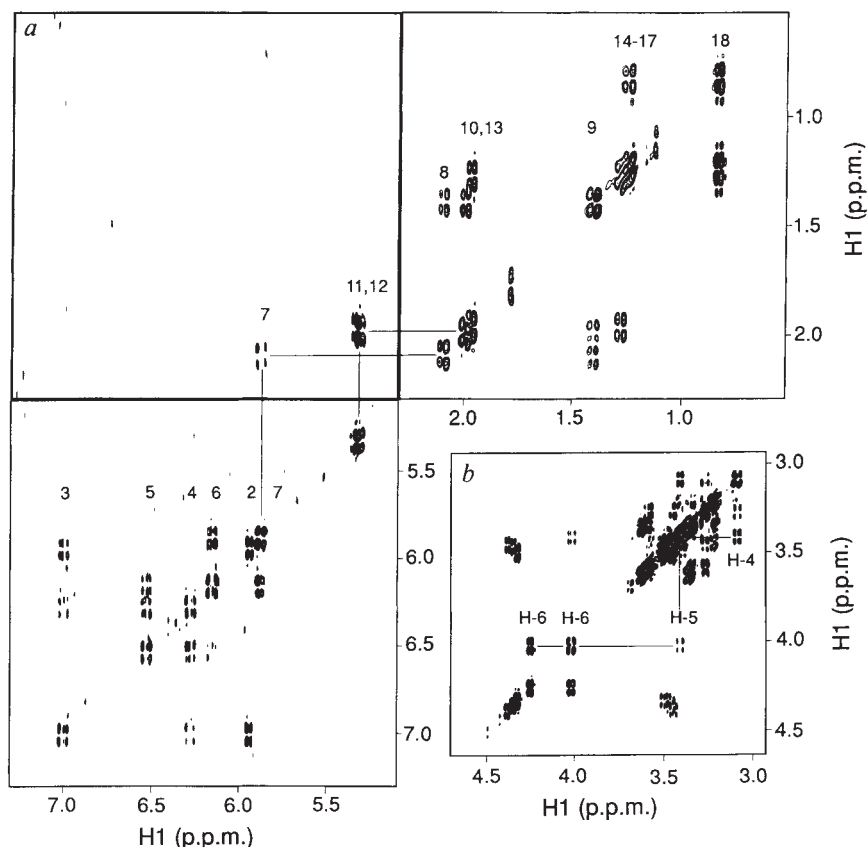


FIG. 4 Structures of Nod metabolites of *R. leguminosarum* bv. *viciae*, a, *nodABCFL*-dependent metabolites; b, *nodABCL*-dependent metabolites. In the nomenclature used, the roman numeral following the species indication (Rlv) refers to the number of glucosamine units and the terms in parenthesis indicate *O*-acetylation, fatty-acyl carbon number and degree of unsaturation, respectively. Also indicated is the predicted nominal M_r of the Nod metabolites.

METHODS. For the large-scale isolation of Nod metabolites, use was made of the strains LPR5045.pIJ1089 (ref. 34) and LPR5045.pIJ1089 *nodE1::Tn5* (ref. 28). Of these strains, *n*-butanol extracts of 200 l and 40 l cultures, respectively, were prepared which were further purified using the methods described in Fig. 1 with the omission of the Sephadex LH20 step. The HPLC elution peaks were further separated on a HPLC PEP-S column (Pharmacia LKB) using isocratic conditions (usually 33/67 or 40/60 acetonitrile/water was used). After this purification procedure about 10–20 mg of each of the wild-type Nod metabolites was obtained. The structure of these Nod metabolites was deduced from the mass spectrometry (Fig. 2), chemical analysis (Fig. 2) and NMR spectrometry data (Fig. 3). The data also show that the wild-type Nod metabolites of $M_r=1,297$ and 1,094 are identical with the corresponding Nod metabolites produced by the *nodE* mutant.

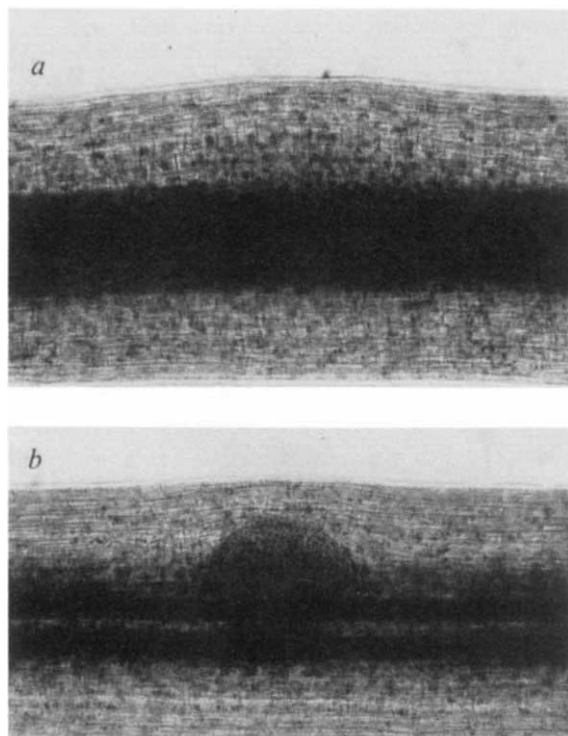
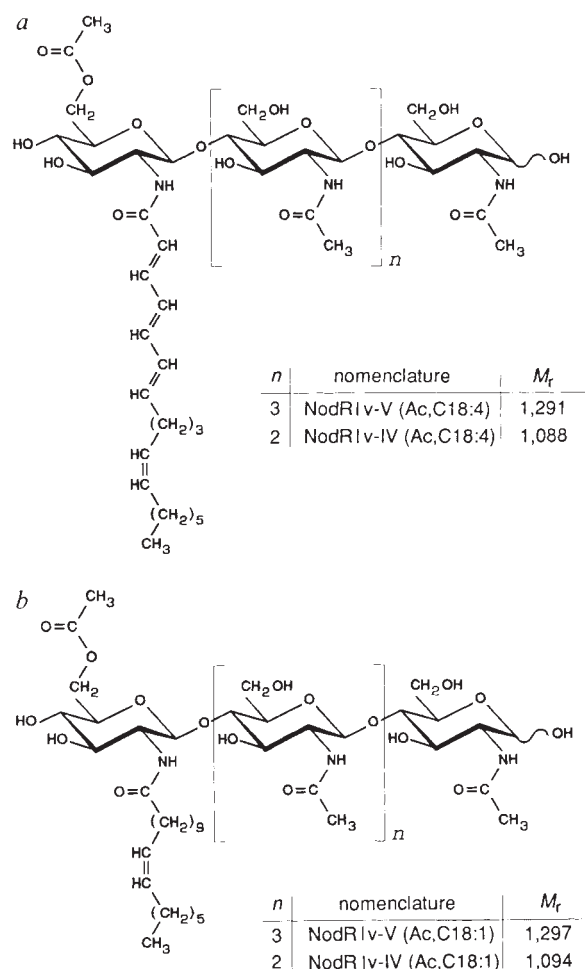


FIG. 5 Nodule meristem (a) induced by NodRlv-V (Ac, C18:4) and a lateral root meristem (b) in a *V. sativa* root. Typical nodule meristems are characterized by a relatively broad base of anticlinically dividing cells in the inner cortex, well distinguishable from the cone-shaped meristem of a young lateral root. After 5 days of growth in the presence of 5×10^{-7} M of the Nod metabolites NodRlv-V (Ac, C18:4) or NodRlv-IV (Ac, C18:4), an average of 12 nodule meristems were induced on each plant. Such nodule meristems were never observed with the other Nod metabolites, also not in the presence of the detergent CHAPS (Table 1). At lower concentrations of the active Nod metabolites less nodule meristems were observed (for example on average two meristems per plant at a concentration of 5×10^{-8} M). The nodule meristems induced by Nod metabolites did not develop further into full-grown root nodules.

METHODS. *V. sativa* plants were cultured on liquid Jensen medium¹⁴ in special glass containers in which the roots were shielded from light. Histological observations were made on whole undissected roots using the method developed by Truchet *et al.*^{12,21}. For each tested concentration of Nod metabolites at least 10 plants were analysed. The molar concentration of stock solutions of NodRlv-V (Ac, C18:4) and NodRlv-V (Ac, C18:1) were measured by weighing 1.0 mg of the dry matter before dissolving it in 50 ml 50/50 acetonitrile/water. The concentration of the other Nod metabolites was estimated by comparison of the integrated peaks of ultraviolet absorbance in HPLC to the standard samples with known concentration of NodRlv-V (Ac, C18:4) in case of the C18:4 containing Nod metabolites or NodRlv-V (Ac, C18:1) in the case of the other Nod metabolites. The Nod metabolites and the chitin fragment (Table 1) were serially diluted in 50/50 acetonitrile/water and 100 μ l samples of the resulting dilutions were added to culture tubes and evaporated to dryness at 80 °C for 10 min.

of oligosaccharide chain length remains unknown. An *O*-acetyl group is also present on a recently discovered new signal compound produced by *R. meliloti*^{10,12}. In this case the *O*-acetyl group did not seem to be essential for the induction of nodule meristems¹². This result indicates that the *O*-acetyl substituent also has host-specific characteristics. Remarkably, the *nodO* gene which encodes an excreted protein which is homologous to haemolysin of *E. coli*^{23,24} is able to complement partially a

deletion of the *nodFEL* genes of *R. leguminosarum*²⁵. This observation indicates that nodulation could occur through either of two pathways encoded by nonhomologous genes. The comparison of the effects of Nod metabolites produced by *nod* mutant strains and those produced by different wild-type rhizobia will be a powerful tool for further study on mechanisms by which Nod signals trigger the process of plant cell division and other root responses. □

Received 19 June; accepted 2 October 1991.

- Long, S. R. *Cell* **56**, 203–214 (1989).
- Spaark, H. P. et al. *EMBO J.* **8**, 2811–2818 (1989).
- Bibb, M. J., Biro, S., Motamedi, H., Collins, J. F. & Hutchinson, C. R. *EMBO J.* **8**, 2727–2736 (1989).
- Shearman, C. A., Rossen, L., Johnston, A. W. B. & Downie, J. A. *EMBO J.* **5**, 647–652 (1986).
- Geiger, O., Spaark, H. P. & Kennedy, E. P. *J. Bact.* **173**, 2872–2878 (1991).
- Downie, J. A. *Molec. Microbiol.* **3**, 1649–1652 (1989).
- Musso, R. E. & Zabin, I. *Biochemistry* **12**, 553–557 (1973).
- Lerouge, P. et al. *Nature* **344**, 781–784 (1990).
- Roche, P. et al. in *Advances in Molecular Genetics of Plant-Microbe Interactions* vol. 1 (eds Henneke, H. & Verma, D.P.S.) 119–126 (Kluwer, Dordrecht, The Netherlands, 1991).
- Roche, P., Lerouge, P., Ponthus, C. & Promé, J. C. *J. biol. Chem.* **266**, 10933–10940 (1991).
- Spaark, H. P. et al. in *Advances in Molecular Genetics of Plant-Microbe Interactions* vol. 1, (eds Henneke, H. & Verma, D.P.S.) 142–149 (Kluwer, Dordrecht, The Netherlands, 1991).
- Truchet, G. et al. *Nature* **351**, 670–673 (1991).
- Spaark, H. P. et al. *Molec. Pl.-Microbe Interact.* (in the press).
- van Brussel, A. A. N. et al. *J. Bact.* **165**, 517–522 (1986).
- van Brussel, A. A. N. et al. *J. Bact.* **172**, 5394–5401 (1990).
- Recourt, K., Schripsema, J., van Brussel, A. A. N., Kijne, J. W. & Lugtenberg, B. J. J. *Pl. molec. Biol.* **16**, 841–852 (1991).
- Recourt, K. thesis, Univ. Leiden (1991).
- Barker, S. A., Foster, A. B., Stacey, M. & Webber, J. M. *J. chem. Soc.* **451**, 2218–2227 (1958).
- Truchet, G., Michel, M. & Dénarié, J. *Differentiation* **16**, 163–173 (1980).
- Finan, T. M. et al. *Cell* **40**, 869–877 (1985).
- Truchet, G., Camut, S., de Billy, F., Odorico, R. & Vasse, J. *Protoplasma* **149**, 82–88 (1989).

- Schwedock, J. & Long, S. R. *Nature* **348**, 644–646 (1990).
- de Maagd, R. A. et al. *J. Bact.* **171**, 6764–6770 (1989).
- Economou, A., Hamilton, W. D. O., Johnston, A. W. B. & Downie, J. A. *EMBO J.* **9**, 349–354 (1990).
- Downie, J. A. & Surin, B. P. *Molec. Gen. Genet.* **222**, 81–86 (1990).
- Hooykaas, P. J. J., Snijderwint, F. G. M. & Schilperoord, R. A. *Plasmid* **8**, 73–82 (1982).
- Spaark, H. P., Okker, R. J. H., Wijnffelman, C. A., Pees, E. & Lugtenberg, B. J. J. *Pl. molec. Biol.* **9**, 27–39 (1987).
- Wijnffelman, C. A., Pees, E., van Brussel, A. A. N., Okker, R. J. H. & Lugtenberg, B. J. J. *Arch. Microbiol.* **143**, 225–232 (1985).
- Sheeley, D. M. & Reinhold, V. N. *Analyt. Biochem.* **193**, 240–247 (1991).
- Hahn, M. G. et al. in *Molecular Plant Pathology: A Practical Approach* (ed. Bowles, D. J.) Chapter 2 (IRL, Oxford, UK, in the press).
- Clucanu, I. & Kerek, F. *Carbohydr. Res.* **131**, 209 (1984).
- Stellner, K., Saito, H. & Hakamori, S. *Arch. Biochem. Biophys.* **155**, 464–472 (1973).
- Breitmaier, E. & Voelter, W. *Carbon-13 NMR Spectroscopy* 3rd edn (VCH, New York, 1987).
- Downie, J. A., Ma, Q.-S., Knight, C. D., Hombrecht, G. & Johnston, A. W. B. *EMBO J.* **2**, 947–952 (1983).

ACKNOWLEDGEMENTS. We thank A. Wijffes (Leiden University) for technical assistance, and C. A. Wijnffelman (Leiden University) and J. Dénarié (CNRS-INRA, Toulouse) for discussions. The research was supported in part by a Department of Energy (DOE) grant as part of the USDA/DOE/NSF Plant Science Centers Program (to the CCRC), by a NSF grant (to V.N.R.), by National Institute of General Medical Sciences grants (to E.P.K.) and by the Royal Netherlands Academy of Arts and Sciences (H.P.S.).

SUPPLEMENTARY INFORMATION. Requests should be addressed to the London editorial office of *Nature*.

LETTERS TO NATURE

Two radio supernovae in the unusual galaxy Markarian 297

Q. F. Yin & D. S. Heeschen

National Radio Astronomy Observatory, Edgemont Road, Charlottesville, Virginia 22903, USA

STARBURST galaxies have higher star formation and supernova rates than normal galaxies, triggered in many cases by the interaction of two galaxies¹. Direct estimates of the supernova rate in starburst galaxies is difficult, because optical direction of supernovae is hampered by high background brightness and by dust², both of which are associated with the starburst phenomenon. Radio observations do not suffer from these problems. We report here the discovery of two radio supernovae in the clumpy irregular galaxy Markarian 297 (Mkn 297, also known as NGC6052), which has been estimated³ to have a star formation rate ten times that of a normal spiral galaxy, and a supernova rate of 0.3–0.5 per year. It may consist of two interacting galaxies⁴. A variable radio source found in 1982 (ref. 5) was attributed to possible supernova activity, and our observations with the Very Large Array, combined with earlier published data, define a reasonable radio light curve for a supernova that exploded in July 1979. The variability resembles that of the young supernova SN1986J (ref. 6), but at maximum the supernova in Mkn 297 would have been about six times as bright. We also identify a second source in the galaxy as a supernova. The presence of two extraordinarily bright supernovae in the same region may be a particular result of the interaction between two galaxies.

A radio image of Mkn 297, derived from our VLA observations, is shown in Fig. 1. There are two principal regions of radio emission, each with a weak tail extending southward. The two emission regions and their tails agree, qualitatively, in position and orientation with the two interacting systems. With the 4-arcsec resolution of this map, the southeast emission region

is resolved and the northwest region partially resolved, while the two tails are resolved along their length. Observations at higher resolution show that the northwest region consists of weak extended emission and two sources smaller than 2.3 arcsec. We call these sources Mkn 297A and B; their 1950 coordinates are, at 20 cm wavelength, 16 h 03 min 01.02 s, 20° 40' 43.6" and 16 h 03 min 01.31 s, 20° 40' 46.8", respectively. Mkn 297A is the previously reported variable⁵. Both sources are well to the north of the centre of the optical image and of the nuclei of the two interacting systems (by at least 2.5 kpc).

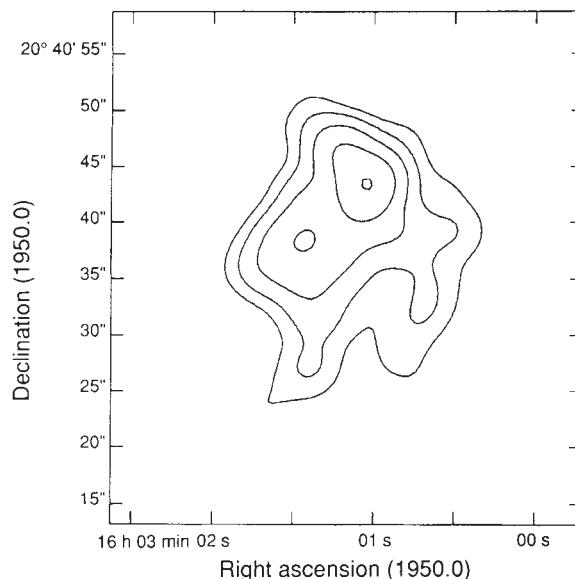


FIG. 1 Contour map of Mkn 297 at 6-cm wavelength, from Very Large Array C-configuration observations made 7 January 1987. Contour intervals are 0.36, 0.72, 1.44, 2.88 and 5.76 mJy per beam. The resolution is 4 arcsec and the r.m.s. noise is $\sim 60 \mu\text{Jy}$ per beam.

TABLE 1 VLA observations of Mkn 297A

Epoch	Configuration	$S_{20\text{cm}}$ (mJy)	$S_{6\text{cm}}$ (mJy)	Ref.
1980 Jan 12	A/B		4.2(0.4)	5
1981 Mar 19	A	6.8 (0.2)		This work
1982 Apr 23	A		12.2 (0.2)	5
1983 Nov 15	A/B	14.0		11
1986 July 27	B	6.4 (0.2)	3.8 (0.1)	This work
1987 Jan 7	C		3.2 (0.1)	This work
1987 Nov 15	B	5.7 (0.7)		See text
1990 July 14	B	4.7 (0.1)		See text
1990 Nov 2	C		2.0 (0.2)	See text

Flux densities of Mkn 297A at different epochs are listed in Table 1. The values in parentheses in columns 3 and 4 are estimated or published uncertainties. The flux density on 1987 November 15 was derived from the visibility data obtained by Condon *et al.*⁷. The values for 1990 are adapted from Q.F.Y. and T. X. Thuan (manuscript in preparation). All of the 1986, 1987 and 1990 flux densities were derived using only baselines longer than 25 thousand wavelengths, to suppress extended emission and facilitate comparison with the higher-resolution observations made with the A and A/B configurations of the Very Large Array. The data are plotted in Fig. 2. We made least-squares fits to the data in Table 1, using Chevalier's model⁸:

$$S_\nu = \text{const} \times \left(\frac{\lambda}{20 \text{ cm}} \right)^{(\gamma-1)/2} \left(\frac{t}{t_{20}} \right)^{-(\gamma+5-6m)/2} \times \exp \left[- \left(\frac{\lambda}{20 \text{ cm}} \right)^2 \left(\frac{t}{t_{20}} \right)^{-3m} \right]$$

where S_ν is the radio flux, λ is the wavelength of observation, γ is the power law index of the electron energy spectrum, t_{20} is the time at which the circumstellar medium exterior to the emitting shell has a free-free absorption optical depth of unity at 20 cm, and m is given by $R_s \propto t^m$ where R_s is the outer shock radius. The best-fit model, also shown in Fig. 2, indicates that the optical event occurred on 1979 July 10 $\pm$ 17.6 days; $t_{20} = 847 \pm 44.2$ days; $\text{const} = 38.4 \pm 2.67$ mJy; $m = 0.8$ and $\gamma = 2.47 \pm 0.13$. The fit shows, at each wavelength, a rapid rise followed by a slow decline. Maximum flux density occurs first at the shorter wavelength, and the spectral index, α , after maximum is ~ -0.7 ($S_\nu \propto \nu^\alpha$). These characteristics are similar to those of young supernovae⁶. With just five observational points at each frequency, the shapes of the radio light curves are not well defined, especially during the rise and around maximum. The solid and dashed lines represent a reasonable fit to the data, however, and are consistent with the radio light curves of other supernovae^{6,8}. We therefore suggest that this is also a supernova.

If Mkn 297A is indeed a radio supernova it is the most powerful yet found. At 6-cm wavelength, the maximum luminos-

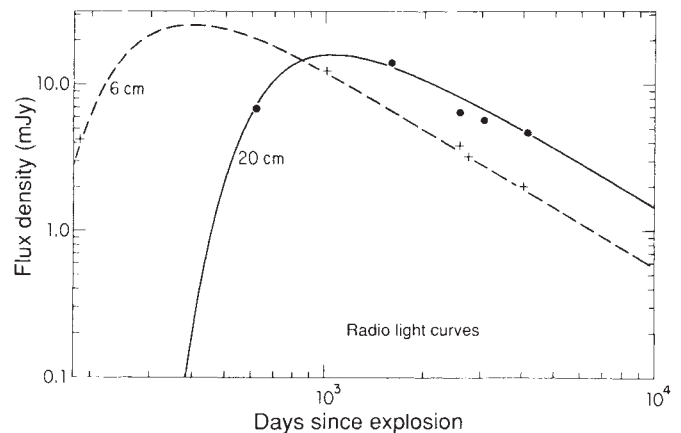


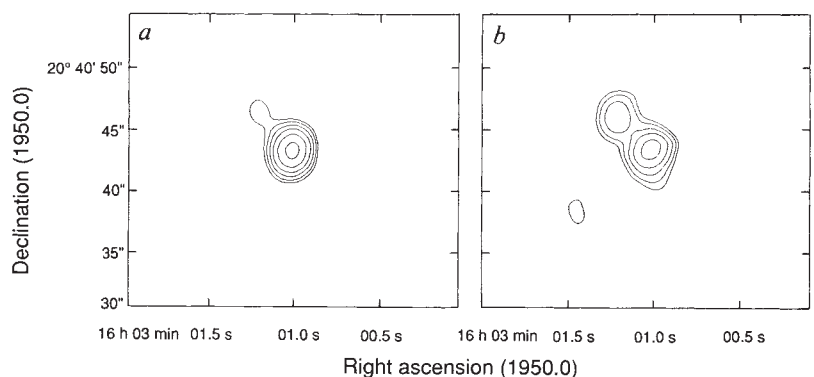
FIG. 2 Radio light curves at 6 and 20 cm. The time $t=0$ corresponds to 10 July 1979, the best-fit date of the supernova explosion.

ity of Mkn 297A, obtained from the best fit, was 1.2×10^{29} erg $\text{s}^{-1} \text{Hz}^{-1}$ (for $v = 4,800 \text{ km s}^{-1}$ (ref. 9) and $H_0 = 75$). This is ~ 6 times as great as SN1986J, 1.4×10^4 times as great as Cas A and 2×10^5 that of SN1987A^{2,6}. Mkn 297A also has a very rapid radio turn-on, ~ 400 days at 6 cm. The high luminosity and rapid turn-on do not fit the correlation between these two parameters previously found⁶ for the radio supernovae.

Figure 3 shows 20-cm maps of Mkn 297A and B at two epochs, July 1986 and July 1990. Mkn 297B is ~ 1.6 kpc (5 arcsec) northeast of Mkn 297A. Its flux density increased from 1 mJy in 1986 to 2.6 mJy in 1990, whereas the flux density of Mkn 297A declined during the same period. We suggest that Mkn 297B is also a supernova. If so, its peak radio luminosity occurred in 1987 or 1988 and was somewhat less than that of Mkn 297A but still greater than SN1986J.

Mkn 297A and B together account for only a small fraction of the total radio luminosity of Mkn 297. Most of the radio luminosity is probably the result of past supernova activity. A population of supernova remnants as those in M82¹⁰ would be undetectable at the distance of Mkn 297, and a single supernova as bright as SN1986J would be marginally detectable if caught at just the right time. The occurrence, within a few years, of what may be the two most luminous known radio supernovae within 1.6 kpc of each other in the same galaxy is remarkable. The occurrence of two supernovae in ~ 6 years is consistent with the previously estimated supernova rate³. That both occurred in the same region and both are more luminous than any previous known radio supernovae may be fortuitous but is more likely to reflect particular properties of this region of the interacting system. The high radio luminosity implies high efficiency in generating magnetic fields and relativistic electrons. Shocks associated with the interaction of two galaxies may have triggered formation of the massive stars that were the progenitors

FIG. 3 Maps at 20 cm of Mkn 297A and B, from Very Large Array B-configuration observations made (a) 27 July 1986 and (b) 14 July 1990. The contour intervals are 1.0, 1.4, 2.0, 2.8, 4.0 and 5.7 mJy per beam. The resolution is $3.0'' \times 2.6''$, and the r.m.s. noise is 0.17 mJy per beam.



of these radio supernovae. Whether the interaction is also responsible for the particular properties of the interstellar medium needed to produce the observed luminosity, or whether those conditions existed beforehand, is unclear. □

Received 15 July; accepted 18 September 1991.

1. Keel, W. C. in *Starbursts and Galaxy Evolution* (eds Thuan, T., Montmerle, T. & Tran Thanh Van, J.) 307–312 (Editions Frontières, Paris, 1987).
2. Chevalier, R. A. in *Massive Stars in Starburst* (eds Leitherer, C., Walborn, N., Heckman, T. & Norman, C. 169–181 (Cambridge University Press, 1990).
3. Klein, U., Heidmann, J., Wiebeinski, R. & Wunderlich, E. *Astr. Astrophys.* **154**, 373–375 (1986).
4. Alloin, D. & Duflo, R. *Astr. Astrophys.* **78**, L5–8 (1979).
5. Heesch, D. S., Heidmann, J. & Yin, Q. F. *Astrophys. J.* **267**, L73–76 (1983).
6. Weiler, K. W., Panagia, N. & Sramek, R. A. *Astrophys. J.* **364**, 611–625 (1990).
7. Condon, J. J., Helou, G., Sanders, D. B. & Soifer, B. T. *Astrophys. J. Suppl.* **73**, 359–400 (1990).
8. Chevalier, R. A. *Ann. N. Y. Acad. Sci.* **422**, 215–232 (1984).
9. Huchra, J. *Astrophys. J. Suppl.* **35**, 171–195 (1977).
10. Kronberg, P. P., Biermann, P., Schwab, F. R. *Astrophys. J.* **291**, 693–707 (1985).
11. Hummel, E., van der Hulst, M. J., Keel, W. C. & Kennicutt, R. C. Jr *Astr. Astrophys. Suppl.* **70**, 517–530 (1987).

ACKNOWLEDGEMENTS. We are grateful to R. A. Chevalier and J. J. Condon for discussions and F. R. Schwab for software help. Q.F.Y. was supported in part by US NSF and the Chinese NSF. The NRAO is operated by Associated Universities, Inc., under cooperative agreement with the NSF.

High-resolution optical observations compared to radio structure of the jet in M87

M. Stiavelli, P. Møller & W. W. Zeilinger

European Southern Observatory, D-8046 Garching bei München, Germany

THE jet in the nearby elliptical galaxy NGC4486 (M87) is a prototype of this kind of nuclear activity¹. High-resolution observations over a wide wavelength range have been used to investigate the propagation of energy along the jet and the radiation mechanism, but these studies have been hampered by the much lower resolution available at optical compared with radio wavelengths. Here we present high-resolution multi-colour images of the jet, obtained using image reconstruction techniques. Morphological features in the jet known from radio observations are seen in our optical images. In particular, we find that the radio-bright ‘knots’ are spatially extended, and that there is significant emission in the regions between the knots. There is also evidence for limb-brightening at a number of positions along the jet. Taken together, our observations suggest that the jet propagates essentially free of losses, generating radio and optical emission at its edges, where interstellar material may be encountered.

Until recently, observations with subarcsecond resolution were obtained solely with photographic plates^{2–4}. Results from CCDs have recently become available^{5,6}, but they were characterized by relatively low angular resolution. On the basis of multi-wavelength studies, it was established that there is a close similarity between the optical and the radio data. But large-scale oscillations of the positions of local maxima of the emission (‘knots’) are seen in the optical images, whereas radio data⁷ indicate a filamentary structure. To investigate these issues further, higher-resolution optical data need to be compared with the radio data.

Our observations are described in Table 1. We obtained images using V, R and I filters at the 2.5-m Nordic Optical Telescope (NOT) equipped with a Tektronix TK512-011 CCD camera (512×512 pixels; scale: 0.20 arcsec pixel⁻¹). These observations were complemented with U, B, V and Gunn-z frames taken at the ESO/MPI 2.2-m telescope using EFOSC2 equipped with a coated Thomson CCD (1,024×1,024 pixels; scale: 0.332 arcsec pixel⁻¹). A detailed description of the data reduction will be given elsewhere (W.W.Z., P.M. and M.S., manuscript in preparation). To determine the properties of the

jet, it was necessary to subtract the contribution of the galaxy which is dominant at longer wavelengths. This was done at the level of $\leq 0.1\%$ accuracy by deriving a smooth model of the galaxy using PLEINPOT⁸. The model was calculated by fitting elliptical isophotes to the luminosity distribution of the galaxy. Multiple gaussian curves were fitted to the nuclear region and the unresolved component was defined as the ‘nucleus’. The model was then subtracted from the images, so that the structure of the jet could be isolated.

To determine the morphology of the jet, we reconstructed our images to improve their angular resolution by using a flux-conserving deconvolution procedure⁹. The reconstruction was performed separately on each image to identify more easily the faint features of the jet. We checked the results of our deconvolution procedure by comparing the V-band frame taken at the ESO/MPI 2.2-m telescope deconvolved to a seeing full width at half maximum (FWHM) of 0.6 arcsec with the original NOT V-band frame. The comparison verified the reliability of our method. In Table 1, a comparison between observed resolution and that obtained after deconvolution is given for each frame. The FWHM was measured on stellar-like images detected on the respective galaxy CCD frames. Their width is consistent with the size of the ‘unresolved’ nuclear component. For the NOT frames, the width in the north–south direction is smaller than that in the east–west direction. In Table 1 we give the average value. Because for the NOT frames the stellar-like images may actually be M87 globular clusters, the values given for the FWHM of the seeing profile are upper limits. All the NOT images were aligned and averaged to increase the signal to noise ratio further. Before alignment, the images were divided into a smaller pixel size (0.067 arcsec pixel⁻¹) to reduce losses in resolution. Figure 1 displays the jet before (panel *a*) and after (panel *b*) deconvolution.

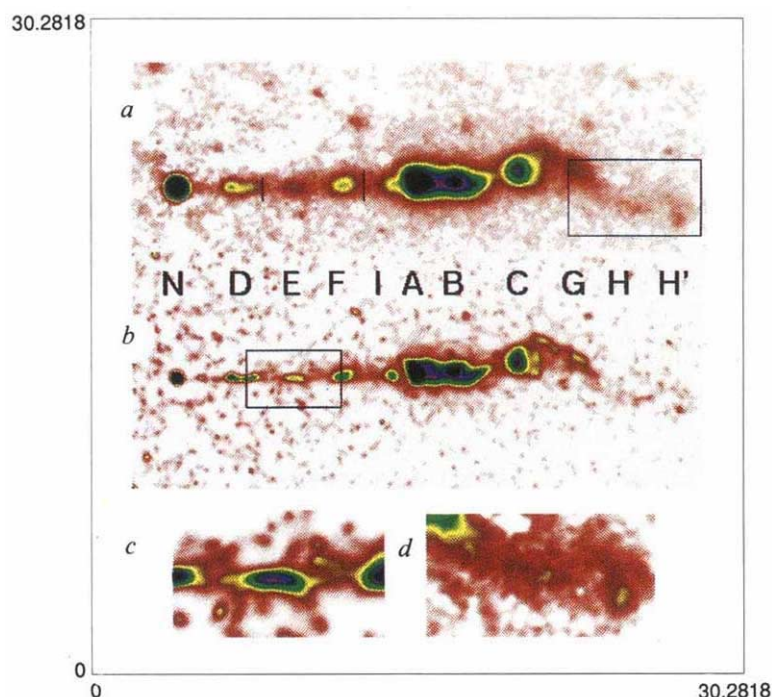
It is readily apparent that all the main knots or knot groups identified previously are present, and that the overall morphology suggested by the Very Large Array radio maps⁷ is seen also at optical wavelengths. The jet is not a mere superposition of discrete sources. At a resolution of 0.25 arcsec, the D, E and F knots and the inter-knot regions clearly contain diffuse emission (Fig. 1c). This is in agreement with the occurrence of filaments in the radio images. The same applies to most of the other knots. The overall morphology seems to be very similar in all colour bands, when the different angular resolution of the various images is taken into account. The presence of a very sharp discontinuity at knot A¹⁰ is confirmed, as is its inclination with respect to the jet axis ($\sim 75^\circ$). The knot complexes D, E and F are arranged in a wavy pattern along the jet axis. These ‘oscillations’ are unrelated to the discontinuity at knot A³ and the jet morphology downstream of A. Instead, they appear because of the presence of a filamentary structure even in the optical.

Analysis of the structure of the inter-knot regions reveals evidence of limb-brightening at several positions along the jet. This phenomenon has been first detected on high-resolution 2-cm maps, and been interpreted as being due to gas filaments wrapped around the jet⁷. It has also been observed in ultraviolet images obtained with the Faint Object Camera of the Hubble Space Telescope (P. Crane, personal communication). The NOT images show this effect in the region of knots D, E and F both before and after deconvolution (Fig. 1c). In Fig. 2, we present two north–south slices perpendicular to the jet axis at the inter-knot regions F–I and D–E which clearly demonstrate the double-horned profile. The slices were obtained from the original (not reconstructed) NOT images. It is evident from Fig. 2 that the effect of limb-brightening is present in all the colours. The effect is less evident in the I-band profile. This may be due to the lower ratio of jet light to galaxy light in the I band, as stars are ‘redder’ than the jet. An additional effect may be caused by alignment errors, as the I image was obtained by adding two frames aligned before subtracting the galaxy model. The sub-

TABLE 1 Observations

Filter	Telescope	Observation date	Exposure time (s)	Seeing FWHM (arcsec)	Deconvolved FWHM (arcsec)
U	ESO/MPI 2.2-m	1991 June 8	1,800	1.35	0.655
B	ESO/MPI 2.2-m	1991 June 8	900	1.37	0.425
V	ESO/MPI 2.2-m	1991 June 8	180	1.22	0.574
V	NOT 2.5-m	1991 May 18	300	0.63	0.277
R	NOT 2.5-m	1991 May 18	300	0.58	0.241
I	NOT 2.5-m	1991 May 18	2 × 300	0.64	0.241
Gunn z	ESO/MPI 2.2-m	1991 June 9	840	1.15	0.472

FIG. 1 Combined frames obtained by averaging all the NOT images before (a) and after (b) deconvolution. The FWHM of a stellar profile is 0.6 in *a* and 0.25 arcsec in *b*. *c*, Enlargement of knot E on the deconvolved image. Diffuse emission and filamentary structure are seen. *d*, The complex of knots H and H' in the original image.



sequent alignment of the galaxy-subtracted I image with the R and V images was performed with higher accuracy on the frames that had been divided into smaller bins.

The nucleus in our deconvolved images appears to be slightly elongated along the E-W direction, in agreement with radio observations⁷. The precise elongation shows variations from colour to colour which are probably due to slight differences in galaxy subtraction and noise, 'amplified' by the reconstruction procedure. We identify a new knot lying between the nucleus and knot D, at a distance of ~ 1.2 arcsec from the former. The possible existence of such a knot was previously suggested in the optical³ and then clearly detected at radio wavelengths⁷ (their knot G). We do not find any evidence for an optical counter-jet. The jet to counter-jet ratio for the entire 20-arcsec-long jet as determined from the combined NOT images is $R \geq 450$ (1σ level, $R \geq 220$ at the 3σ level).

The structure of the jet downstream of A is more complex. In addition to the sharp discontinuity at A, we observe filamentary structure and limb brightening. We have attempted to determine whether colour gradients are present in the A-B-C knot complex. Because the knots are resolved in the NOT images (FWHM of A is ~ 1.5 arcsec, as compared with the seeing, which

was ~ 0.6 arcsec), their resolution is marginally adequate to detect any such gradient. Such a study is complicated, however, by the presence of steep gradients in the fluxes, which may produce large spurious gradients in case of even small misalignments. It seems that the centre of the emission in knots A, B and C is redder than the jet edges. The effect should, however, be confirmed by higher-resolution observations performed with the Hubble Space Telescope. An analysis of the spectral index variations along the jet shows a steepening outward of knot A,

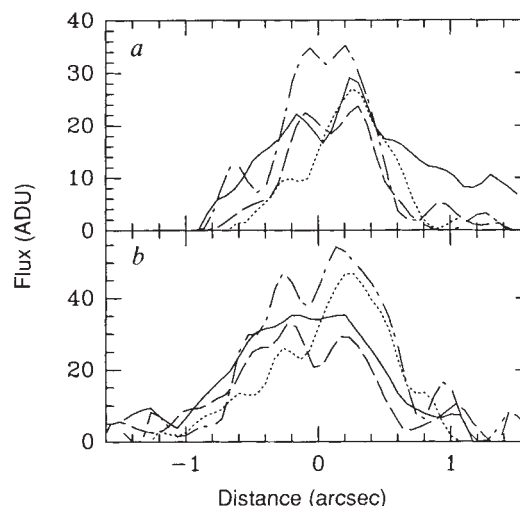


FIG. 2 Slices perpendicular to the jet axis obtained at the inter-knot regions F-I (a) and D-E (b). The slices are extracted with a 0.33 arcsec slit from the NOT V (---), R(---), I(····) and the combined NOT frames of Fig. 1 (—). The flux units (ADU) are arbitrary. The limb-brightening discovered in radio maps is here seen also in the optical.

in agreement with other authors^{1,5}. Details will be published elsewhere (W.W.Z., P.M. and M.S., manuscript in preparation).

Our main conclusions can be summarized as follows: the existence of an inner knot is confirmed; extended emission is present in the knots, which are not simple superpositions of discrete sources; evidence for limb brightening is also seen in the optical band.

The physical picture of a jet where energy propagates outwards with little loss until external material is encountered at knot A thus seems essentially confirmed. Observations are consistent with a cone-like and hollow jet, with emission occurring mostly at its edges and along filaments which occasionally become brighter. If a dependence of the spectral index on the position perpendicular to the jet is confirmed, then this would show that the emission mechanism at the 'shock-dominated' knots (A and outward) is different from that closer to the nucleus. □

Received 10 July; accepted 16 September 1991.

1. Biretta, J. A., Stern, C. P. & Harris, D. E. *Astr. J.* **101**, 1632–1646 (1991).
2. Nieto, J.-L. & Lelièvre, G. *Astr. Astrophys.* **109**, 95–100 (1982).
3. Lorre, J. J. & Nieto, J.-L. *Astr. Astrophys.* **130**, 167–170 (1984).
4. Lelièvre, G., Nieto, J.-L., Horville, D., Renard, L. & Servan, B. *Astr. Astrophys.* **138**, 49–56 (1984).
5. Pérez-Fournon, I., Colina, L., González-Serrano, J. I. & Biermann, P. L. *Astrophys. J.* **329**, L81–L84 (1988).
6. Keel, W. *Astrophys. J.* **329**, 532–550 (1988).
7. Owen, F. N., Hardee, P. E. & Cornwell, T. J. *Astrophys. J.* **340**, 698–707 (1989).
8. Prugniel, P. in *1st ESO-ST-ECF Data Analysis Workshop* (ed. Grosbøl, P. J. et al.) 161–167 (ESO, Garching, 1989).
9. Lucy, L. B. & Baade, D. in *1st ESO-ST-ECF Data Analysis Workshop* (ed. Grosbøl, P. J. et al.) 219–223 (ESO, Garching, 1989).
10. Biretta, J. A., Owen, F. N. & Hardee, P. E. *Astrophys. J.* **274**, L27–L30 (1983).

ACKNOWLEDGEMENTS. We thank J. Biretta for criticisms of the manuscript. The observations were collected at the NOT, La Palma and ESO, La Silla.

Magnetically controlled convection in a paramagnetic fluid

D. Braithwaite, E. Beaugnon & R. Tournier

Centre de Recherches sur les Très Basses Températures, CNRS,
BP 166, 38042 Grenoble Cedex 9, France

CONVECTION in a liquid is important for problems involving heat transfer and crystal growth from a melt. The driving force for convection is usually the density difference between hot and cold regions of the fluid. If the fluid has a magnetic susceptibility that varies with temperature, magnetic forces, rather than buoyancy, can be made to drive convective motion. Studies on ferrofluids (suspensions of ferromagnetic particles¹) have shown that magnetic convection can be initiated in a homogeneous magnetic field^{2,3} and enhanced in a field gradient⁴. We show here that the strong magnetic fields available from superconducting magnets can be used to induce magnetic convection in normal paramagnetic fluids, such as solutions of paramagnetic salts or melts of paramagnetic solids. We have used a magnetic field both to enhance and to suppress buoyancy-driven convection in a solution of gadolinium nitrate, the sign of the effect depending on the relative orientation of magnetic-field and temperature gradients. The effect might be exploited in heat-transfer devices or to control microstructures in crystal growth.

For a paramagnetic fluid, the magnetization, M , is proportional to the applied field: $M = \chi H$. Here χ is the magnetic susceptibility, the temperature dependence of which follows Curie law behaviour ($\chi \propto T^{-1}$). In a magnetic field gradient, the force per unit volume on the fluid is

$$F = \mu_0 M \frac{\partial H}{\partial z} = \mu_0 \chi H \frac{\partial H}{\partial z}$$

The important quantity for the magnetic force is therefore the

square field gradient. We will express numerical values of $\mu_0^2 H(\partial H/\partial z)$ in units of $T^2 m^{-1}$.

Convection in the fluid increases the efficiency of heat transfer and is quantified by the Nusselt number, the ratio of the apparent thermal conductivity in the presence of convection to that in its absence. In the absence of a magnetic field, buoyancy-driven convection is governed by the Rayleigh number $Ra = g\beta\Delta T\delta^3/\alpha\nu$, where g is the acceleration due to gravity, β is the coefficient of thermal expansion, δ is the vertical height of fluid in the Bénard configuration, ΔT is the temperature difference, α is the thermal diffusivity and ν the kinematic viscosity.

For a paramagnetic fluid in a magnetic field gradient, the magnetic equivalent of β is

$$-\frac{1}{\chi} \frac{\partial \chi}{\partial T} \cong \frac{1}{T}$$

and the term $g\beta$ is replaced by $\beta(g + F_m) + F_m T^{-1}$, where

$$F_m = \frac{\mu_0 \chi}{\rho} H \frac{\partial H}{\partial z}$$

is an acceleration (force per unit mass) due to the magnetic force on the fluid, which can be positive or negative, and ρ is the density of the liquid. We can therefore define a magnetic Rayleigh number

$$Ra_m = Ra(1 + F_m g^{-1}(1 + (T\beta)^{-1})) \quad (1)$$

The magnetic part becomes dominant when $|F_m g^{-1}(1 + (T\beta)^{-1})| > 1$. As the field gradient is independent of the temperature gradient, the magnetic and buoyant contributions may be of the same sign or of opposite sign. Convection can be either enhanced or reduced.

The paramagnetic fluid, prepared by dissolving 4.33 g of gadolinium nitrate in 10 ml of water, has a measured specific susceptibility of $1.63 \times 10^{-7} m^3 kg^{-1}$ at room temperature. The magnetic term will therefore become dominant for a square field

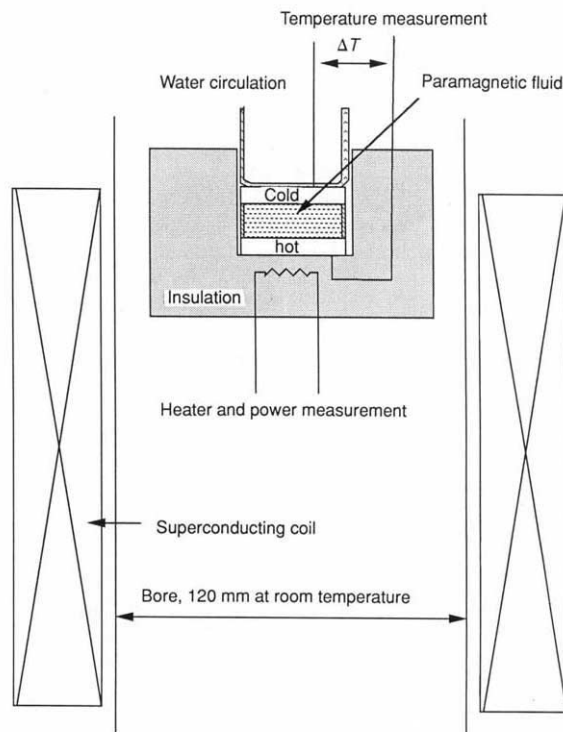


FIG. 1 Schematic view of the experimental setup in the superconducting coil. The convection cell can be used with the heater on the top or the bottom, and can be placed either above or below the coil centre.

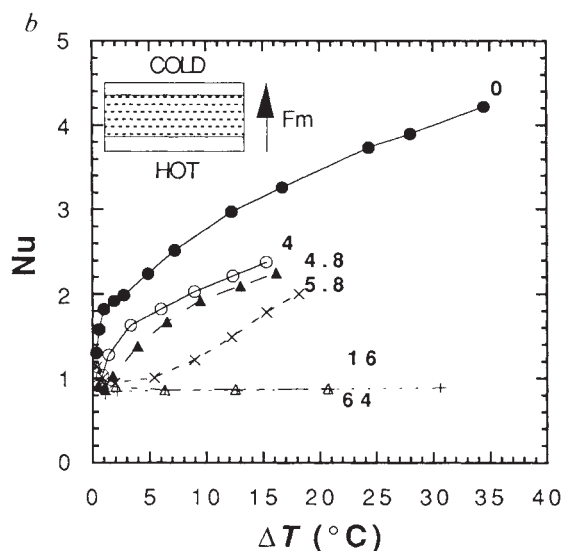
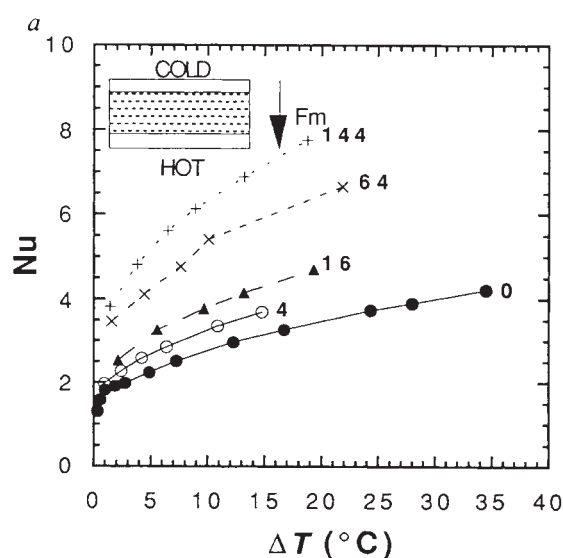


FIG. 2 Nusselt number as a function of temperature difference for configurations 1 and 2. Numbers represent the value of $\mu_0^2 H(dH/dz)$, in units of $T^2 m^{-1}$, for each curve, and the insets show the experimental configuration.

a, Configuration 1; b, configuration 2. In the latter case, natural convection is inhibited by the field, and is suppressed totally for $\mu_0^2 H(dH/dz) > 15 T^2 m^{-1}$.

gradient of $5 T^2 m^{-1}$. The fluid was placed in a 10-mm-high cylindrical cell (diameter 25 mm) closed with two horizontal plates (Fig. 1). One plate was kept at constant temperature with flowing water while the other was heated by a constantan heater. The power was monitored by measuring the current and voltage. The temperature gradient was measured with a differential copper-constantan thermocouple with one junction on each plate. The cell, surrounded by polystyrene insulation, was placed in an 8-T superconducting magnet with a maximum $H dH/dz$ of $250 T^2 m^{-1}$ (commercially available coils could produce a field almost ten times this value). The magnetic force was downwards if the cell was placed above the centre of the coil, upwards if it was below the centre. The heated plate could be either on the top or the bottom. The apparent thermal conductivity, k_{ap} , was measured by monitoring the stabilized temperature gradient between the hot and cold plates for each value of power and magnetic field. The normal thermal conductivity, $k = 0.85 W^\circ C^{-1} m^{-1}$, was measured with the cell heated from above in zero field, and the Nusselt number, $Nu = k_{ap}/k$, was determined for each point.

We used three configurations for the experiments. (1) The cell was heated from below and the magnetic force applied downwards: in this position, natural convection exists and is enhanced by the magnetic contribution. The Nusselt number increases markedly (Fig. 2a) for a given temperature gradient. With the maximum field gradient, the apparent conductivity reached $\sim 7 W^\circ C^{-1} m^{-1}$ for $\Delta T = 20^\circ C$. In contrast, when we performed the same experiment with a ferrofluid (EMG 901 173A), the initial increase in convection quickly reached saturation as the field was increased. The combination of this saturation and the low normal thermal conductivity measured for the ferrofluid ($k = 0.3 W^\circ C^{-1} m^{-1}$) causes heat transfer to be less efficient than for the gadolinium nitrate solution. For $\Delta T = 20^\circ C$, the apparent conductivity did not exceed $2 W^\circ C^{-1} m^{-1}$.

(2) The cell was heated from below and the magnetic force applied upwards: in this case natural convection exists but is inhibited by the applied field (Fig. 2b). For square field gradients greater than $15 T^2 m^{-1}$, convection is suppressed altogether, and a state of pure thermal conduction is reached.

(3) The cell was heated from above and the magnetic force applied upwards: in this case there is no natural convection, but convection can be induced by applying a field. For gradients

above $15 T^2 m^{-1}$, the magnetic effect dominates and the situation is almost identical to case (1).

Although we do not have accurate values for parameters β and ν of the gadolinium nitrate solution, we have used those for water to estimate the magnetic Rayleigh number from equation (1). For large Ra_m , for which convection is enhanced by the field, a plot of Nu against Ra_m is a smooth curve (Fig. 3) similar to that obtained for natural convection. Prediction of the results is therefore possible, provided the essential parameters β , χ and $d\chi/dT$ are approximately known. When the magnetic force opposes convection, on the other hand, the low magnetic Rayleigh number is the result of the difference of two large numbers, and the uncertainties in the fluid parameters make accurate prediction of results difficult.

As the value of $\sim 5 T^2 m^{-1}$ necessary for magnetic effects to dominate thermal effects is obtainable with permanent magnets, and as the efficiency of heat transfer is considerably higher than that measured for a ferrofluid, parafluids such as rare earth salt solutions could prove an alternative to ferrofluids for heat transfer devices. Another possible application is in crystal growth. Reducing convection in a molten liquid during solidification improves the quality and homogeneity of the material; for this

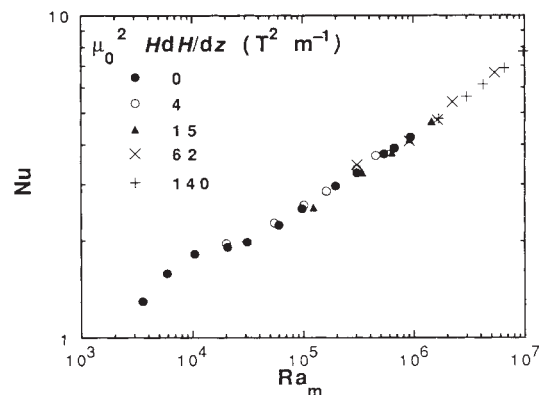


FIG. 3 Global plot of Nusselt against Ra_m for all experiments at different fields. All points lie on a curve characteristic of natural convection. The change of slope around $Ra_m = 10^4$ is due to the transition from the laminar regime to the turbulent regime.

reason, microgravity experiments have been carried out in space⁵. On Earth, a homogeneous magnetic field can reduce convection in conducting fluid⁶. In our experiments, the fluid need not be electrically conducting; provided the susceptibility varies with temperature, we can increase or damp convection by placing the fluid in a magnetic field gradient. In addition to producing the levitation⁷ and orientation⁸ effects described previously, it may be possible to use magnetic fields to control the growth rate and microstructure of materials. □

Received 10 July; accepted 24 September 1991

1. Rosensweig, R. E. *Ferrohydrodynamics* (Cambridge University Press, 1985).
2. Finlayson, B. A. *J. Fluid Mech.* **40**, 753 (1970).
3. Schwab, L., Hildebrandt, U. & Stierstadt, K. *J. Magnet. magn. Mat.* **39**, 113–114 (1983).
4. Nakatsuka K., Hama, Y. & Takahashi, J. *J. Magnet. magn. Mat.* **85**, 207–209 (1990).
5. Carruthers, J. R. *J. Cryst. Growth* **42**, 379–385 (1977).
6. Oreper, G. M. & Szekeley, J. *J. Cryst. Growth* **64**, 505–515 (1983).
7. Beaupre, E. & Tournier, R. *Nature* **349**, 470 (1991).
8. De Rango, P. *et al. Nature* **349**, 770–772 (1991).

ACKNOWLEDGEMENTS. We thank J. Blanchard from Thomson TCMS for advice and M. R. Lees for reading the manuscript. This work was supported by FRAMATOME.

Seasonal and depth-related changes in the source of sinking particles in the North Atlantic

Mark A. Altabet*, Werner G. Deuser*, Susumu Honjo* & Christian Stienen†

* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 USA

† Institut für Meereskunde, Düsterbrook Weg 20, D-2300 Kiel, Germany

LARGE, fast-sinking particles are important in the downward transport and redistribution of biogeochemical species in the deep ocean. Using nitrogen isotope ratio, $^{15}\text{N}/^{14}\text{N}$, as an *in situ* tracer, we investigate the source and transformation of these particles in the North Atlantic ocean. We observe seasonal variations in $\delta^{15}\text{N}$ associated with seasonal changes in near-surface nitrate concentration and particle flux; the nitrogen isotope variations are consistent with, but much larger than, previously observed variability^{1,2}. Our results show that the signal from these near-surface changes propagates rapidly into the deep ocean, but is modified depending on the phase of the seasonal production cycle. Surprisingly, we find that $\delta^{15}\text{N}$ values for sinking particles decrease with depth during low-flux periods—behaviour that may occur generally in the open ocean. The sinking particles must therefore be either gaining light nitrogen or losing heavy nitrogen, an effect that we believe requires there to be another source of sinking particles, apart from recent surface production.

Deeply sinking particles were collected using moored time-series sediment traps at the temperate site of the Joint Global Ocean Flux Study, North Atlantic Bloom Experiment (JGOFS NABE; 47° N, 20° W) and at the subtropical JGOFS time-series station off Bermuda (OFP; 31° 50' N, 64° 10' W). Collection intervals were two weeks at the former and two months at the latter, and trap design and sample handling were as described previously^{3,4}. At both sites, near-surface sinking particles were also collected using floating sediment traps with one-day deployments⁵ (B. von Bodungen *et al.*, manuscript in preparation). Suspended particles were collected by filtration of water samples through glass-fibre filters⁵. Preparation and analysis for nitrogen isotopic ratio included Dumas combustion in sealed ampoules as described previously^{5,6}. Analytical reproducibility was $\pm 0.1\%$.

Seasonal changes in $\delta^{15}\text{N}$ (measured relative to atmospheric N_2 , in ‰ units) have previously been observed in association with spring bloom events^{1,2,7}, but near-surface suspended and sinking particles at the NABE site show the largest and most marked temporal shift in $\delta^{15}\text{N}$ reported so far; values increased by 7‰ over a 25-day period (Fig. 1). These changes in $\delta^{15}\text{N}$ were closely coupled to the corresponding depletion of NO_3^- as the spring bloom progressed and are best explained by isotope fractionation produced during NO_3^- uptake by phytoplankton⁸. Correlation between low $\delta^{15}\text{N}$ values for particulate nitrogen and high surface NO_3^- concentrations has previously been cited as evidence for this effect in the open ocean^{1,9,10}, and our more extensive time-series data are consistent with one-step Raleigh fractionation kinetics associated with NO_3^- assimilation. The estimated fractionation factor of 8–9‰ (see Fig. 1) is well within the range of previous culture studies for NO_3^- uptake⁸.

This large temporal change in near-surface $\delta^{15}\text{N}$ provides a strong source signal for studying the pathways of nitrogen in the open ocean on timescales of weeks to months, timescales which are relevant to particle residence time and fluxes¹¹. In 1989 and 1990, $\delta^{15}\text{N}$ minima and particle flux maxima in the deeper water column were closely associated (Fig. 2). This is evidence that the surface $\delta^{15}\text{N}$ signal propagates downward to between 1,000 and 3,700 m with a time lag of two to three weeks. During the bloom period, there seems to be rapid downward transfer of unaltered surface production (Figs 1 and 2). For the other segments of the time series, however, the $\delta^{15}\text{N}$ data indicate substantial modification in the deep water column. First, deeply

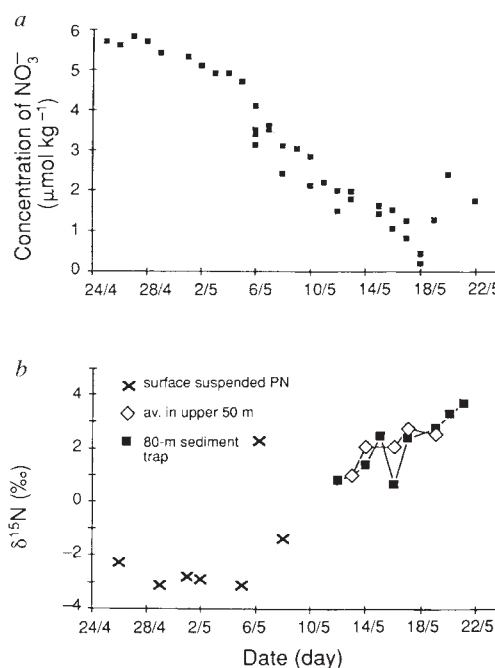


FIG. 1 *a*, Trend with time of NO_3^- concentration at 10 m at the site of the JGOFS NABE (47° N, 20° W) in spring 1989. Up to 6 May, data are from samples collected by the RV *Atlantis II* (JGOFS NABE database). Later samples were collected by the RV *Meteor*. *b*, Trend with time of $\delta^{15}\text{N}$ in surface suspended particles (x: collected by D. Repeta on the RV *Atlantis II*), suspended particles averaged over the upper 50 m (◇), and sinking particles collected by sediment traps (one-day deployments) placed at 80 m (■). Equivalent $\delta^{15}\text{N}$ values for near-surface suspended and sinking particles show that, during the bloom, the processes transforming suspended particles into sinking ones did not alter $\delta^{15}\text{N}$ values. This contrasts with findings for oligotrophic regions⁵. The fractionation factor for NO_3^- uptake can thus be readily estimated by assuming that at the beginning of the observation period, before depletion takes place, $\delta^{15}\text{N}$ for surface NO_3^- is the same as for deep water NO_3^- (5–6‰)³⁰. Accordingly, the fractionation factor (ϵ) is given as the difference in $\delta^{15}\text{N}$ between the source NO_3^- and the product particulate nitrogen (–3‰) at this time, or 8–9‰.

sinking particles are distinct from those near the surface in both $\delta^{15}\text{N}$ value (3‰ higher) and its temporal trend during the onset of the bloom in April and May 1989 (Fig. 2). Second, $\delta^{15}\text{N}$ decreases with depth throughout summer, autumn and winter between 1,000 and 3,700 m. Particle decomposition would have been expected to increase $\delta^{15}\text{N}$ with depth^{5,12}. Third, at any given depth, the highest $\delta^{15}\text{N}$ values are observed in winter, whereas increases in surface NO_3^- due to cooling and destratification should instead have produced a substantial lowering of $\delta^{15}\text{N}$ values.

Observations at the OFP site for moored sediment traps are qualitatively similar to those at the NABE site (Fig. 3). Values of $\delta^{15}\text{N}$ tend to be inversely correlated with particle flux¹. At this site, however, elevated surface NO_3^- concentrations are present for only a few months during late winter and early spring, and they correspond in phase and duration with the minimum in $\delta^{15}\text{N}$. Wintertime primary production at the subtropical OFP site is never severely light-limited, and particle flux maxima in late winter to early spring are triggered by deep convective mixing and subsequent increase in near-surface NO_3^- (ref. 13). Because of insufficient light, wintertime productivity is low at the temperate NABE site despite high surface nutrient concentrations, and this probably accounts for the observation of high $\delta^{15}\text{N}$ at this time. Wintertime production of ^{15}N depleted material is apparently insufficient as a source for deeply sinking particles. Possible high $\delta^{15}\text{N}$ sources include overwintering zooplankton populations¹⁴ ($\delta^{15}\text{N}$ increases substantially with trophic level¹⁵⁻¹⁷) and suspended particles produced in the upper water during the previous summer. In either case, autumn and winter particle flux at the NABE site is strongly decoupled from surface production. At OFP, coupling is reasonably strong throughout the year¹⁸.

At both sites, $\delta^{15}\text{N}$ in sinking particles surprisingly decreases with depth, particularly when flux is low. This behaviour contrasts with the well known increase in $\delta^{15}\text{N}$ with depth observed for suspended particles in the deep ocean^{5,12,19}. At OFP, comparison of $\delta^{15}\text{N}$ values for deep sinking and suspended particles sharply illustrates the differences between these two particle types; between the surface and 3,200 m, $\delta^{15}\text{N}$ in suspended particles increases from near zero to ~7‰, while $\delta^{15}\text{N}$ in sinking

particles decreases from 4‰ to 1‰ (Fig. 4). Similar observations have been made previously in Antarctic waters¹⁶ and the North-west Pacific¹⁹. The small size and limited time coverage of these data sets, in addition to problems of sample integrity in sediment traps¹⁸, however, prevented conclusions being drawn about the generality of these observations. Our more extensive data set for the two North Atlantic sites, together with similar but less complete findings for the Norwegian Basin (M. Voss and M.A.A., unpublished data) and the Santa Monica Basin, California (P. M. Williams *et al.*, manuscript in preparation), indicates that a decrease in $\delta^{15}\text{N}$ values with depth for sinking particles is a ubiquitous phenomenon and, especially when compared with the isotope behaviour of suspended particles, an indicator for important and widespread processes governing particle dynamics in the deep ocean.

The increase in $\delta^{15}\text{N}$ for suspended particles with depth is thought to result from progressive decomposition and the removal of dissolved nitrogen depleted in ^{15}N . Kinetic isotope fractionation, as a rule, leads to enrichment of the heavier isotope in the source relative to product pools, and there is evidence for this effect in protein hydrolysis (S. Macko, personal communication) and deamination reactions²⁰. Similarly, $\delta^{15}\text{N}$ values for sinking particles would be expected to increase with depth, although less than for suspended particles because the relatively rapid sinking speeds allow much less time for decay. Alternatively, extensive interconversion of sinking and suspended particles through disaggregation and reaggregation²¹, which would slow the effective descent of material through the water column, would also homogenize the $\delta^{15}\text{N}$ signal between the two particle types. Instead, our observations indicate, first, that sinking and suspended particles do not extensively interact in the ocean as previously proposed, and second, that in contrast to suspended particles, sinking particles either lose a nitrogen fraction enriched in ^{15}N or gain a fraction depleted in ^{15}N with depth. For the special case where an increase in particle flux accompanies the decrease in $\delta^{15}\text{N}$ with depth (summer 1989 at the NABE site), a third possibility is the horizontal transport of bloom-produced particles from a more productive site²² (M.A.A. and C.S., manuscript in preparation).

In some microbe culture and terrestrial plant material, protein

FIG. 2 *a*, Time course for $\delta^{15}\text{N}$ values for moored sediment-trap collections (<1-mm fraction) at 1,000, 2,000 and 3,700 m depth at the NABE site during and after surface observations. A second spring bloom was evidently sampled in March/April 1990. Points are plotted at the mid-points of the two-week collection intervals, and the envelope of near-surface $\delta^{15}\text{N}$ values (shaded region) is shown for comparison. *b*, Corresponding time course for particulate nitrogen flux. The observed decrease in $\delta^{15}\text{N}$ with depth for the sediment trap samples is unlikely to have been produced by sampling artefacts. Isotope analyses did not include any filter material which might have adsorbed dissolved nitrogen. No alteration of nitrogen content of the collected material while residing in the trap seems to have occurred, as there was no change in total dissolved nitrogen concentration between the ambient water added to the trap cups before and after trap retrieval. Any bias introduced by the cup environment would be likely to increase with time, but the samples collected during the 1989 spring bloom which resided in the cups the longest show the smallest change in $\delta^{15}\text{N}$ with depth.

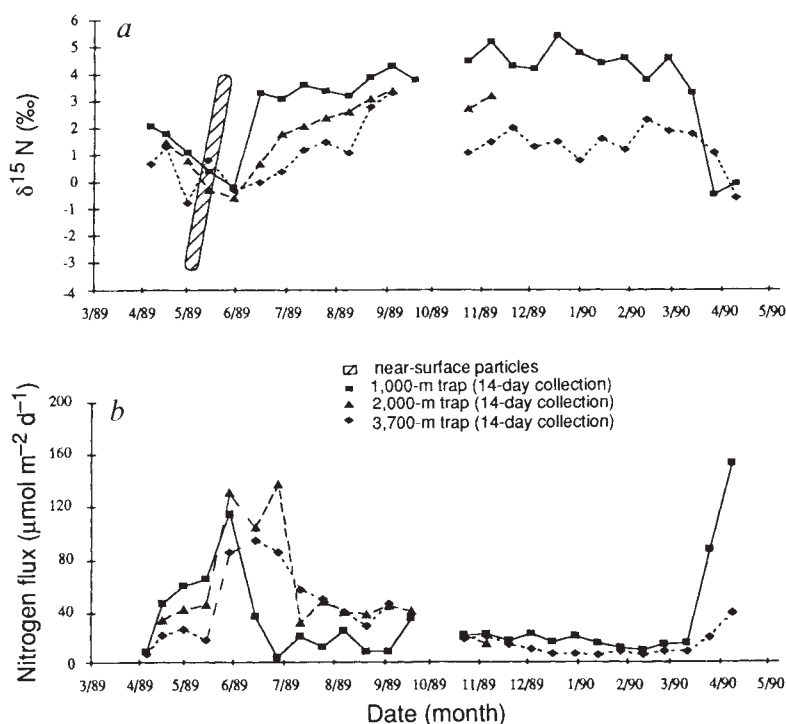
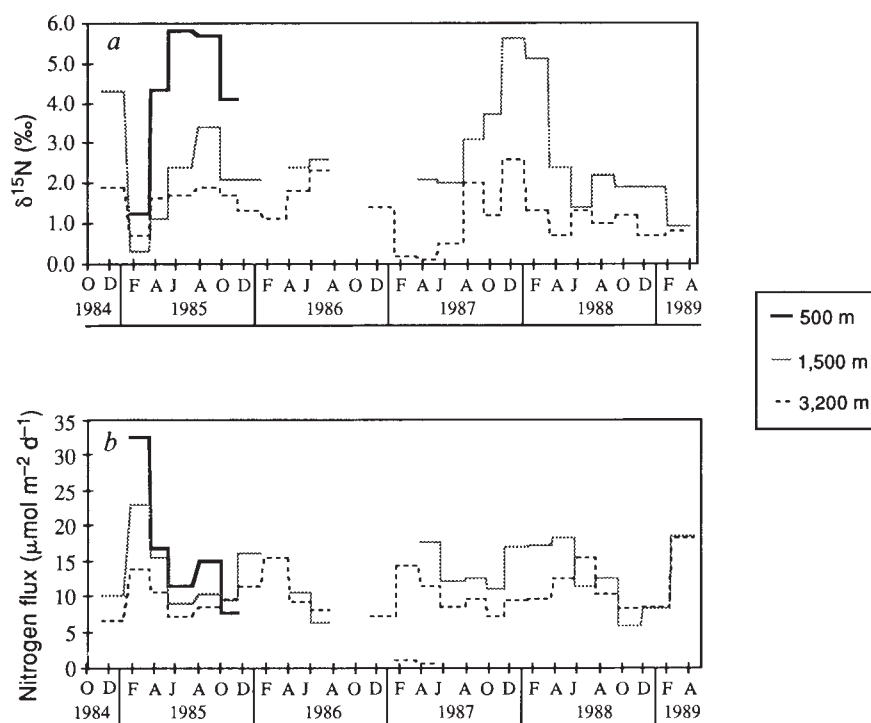


FIG. 3 *a*, Time course (at 2-month intervals) for $\delta^{15}\text{N}$ values for moored sediment-trap collections (<37- μm fraction) at 500, 1,500 and 3,200 m depth at the JGOFS time-series station off Bermuda (OFP; 31° 50' N, 64° 10' W). The horizontal portions of the traces correspond to the roughly two-month collection interval. *b*, Corresponding time course for particulate nitrogen flux.



nitrogen has been found to be 3‰ enriched in ^{15}N relative to total nitrogen²³, and it is possible that protein is selectively removed from sinking particles. This explanation is consistent with the amplification of the decrease in $\delta^{15}\text{N}$ with depth under low flux conditions, which could be due to increased decomposition of amino acids with depth during these periods²⁴. Available data for sediment trap samples collected near the OFP site at 3,200 m do show that total amino acids make a smaller contribution to total nitrogen than in fresh phytoplankton or zooplankton²⁵. But the total amino acid contribution to total nitrogen in

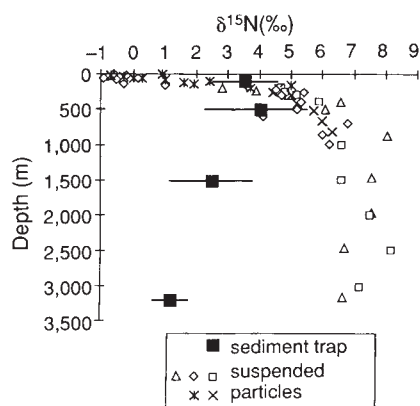


FIG. 4 Comparison of $\delta^{15}\text{N}$ values for material collected in sediment traps and suspended particles filtered from water samples at the OFP site. Means ± 1 s.d. for the sediment-trap data (plotted in Fig. 3) are shown. The 100-m point is the average for seasonally distributed one-day deployments of floating sediment traps ($n=20$). The suspended particle data are from several station occupations in different seasons marked by different symbols. In contrast to the sediment trap samples (≥ 500 m), pre-combusted glass-fibre filters used to collect suspended particles were routinely included in the $\delta^{15}\text{N}$ analysis. Filter and handling blanks have been shown, however, to be very small compared with typical sample sizes³¹. Moreover, equivalent results for $\delta^{15}\text{N}$ and particulate nitrogen concentration in samples collected with conventional bottle samples and *in situ* pumps (which load much larger quantities of particulate matter onto filters) indicate no artefacts associated with adsorption of dissolved nitrogen⁵.

peak flux periods was not significantly different from the time-series average (averages are 39.5 and 42.8‰, respectively).

A low-level, deep source of ^{15}N -depleted material to sinking particles would also produce the greatest signal during low flux periods when the contribution from surface production is smallest. The geographical range of observations suggests that such a source would, moreover, have to be widely distributed in the ocean. Dissolved organic matter has been suggested as a source for deeply sinking particles on the basis of radiocarbon data²⁶, but isotope analysis is needed to determine whether this material is sufficiently depleted in ^{15}N to account for our results. Incorporation of abundant dissolved NO_3^- with associated isotope fractionation by bacteria on sinking particles is another alternative. Such activity may be required to compensate for the organic C/N ratios of sinking particles²⁷, which are high compared with bacterial requirements^{28,29}.

Regardless of mechanism, transformation of fast-sinking particles in the deep ocean is influenced by processes distinct from those affecting the more abundant suspended particle pool. Size is a principal distinction between these particle classes, and micro-environments within large, sinking particles may enhance microbial activity that leads either to the select removal of ^{15}N -enriched material or the addition of ^{15}N -depleted material. □

Received 8 March; accepted 10 September 1991.

- Altabet, M. A. & Deuser, W. G. *Nature* **315**, 218–219 (1985).
- Altabet, M. A. & McCarthy, J. J. *Deep-Sea Res.* **32**, 755–772 (1985).
- Honjo, S. & Doherty, K. W. *Deep-Sea Res.* **29**, 133–149 (1987).
- Deuser, W. G. *Deep-Sea Res.* **33**, 225–246 (1986).
- Altabet, M. A. *Deep-Sea Res.* **35**, 535–554 (1988).
- Minagawa, M., Winter, D. A. & Kaplan, I. R. *Analyt. Chem.* **56**, 1859–1861 (1985).
- Goering, J., Alexander, V. & Haubenstock, N. *Estuar. Coast. Shelf Sci.* **30**, 239–260 (1990).
- Wada, E. & Hattori, A. *Geochim. Cosmochim. Acta* **40**, 249–251 (1976).
- Miyake, Y. & Wada, E. *Rec. Oceanogr. Works Japan* **9**, 37–53 (1967).
- Wada, E. & Hattori, A. *Geochim. Cosmochim. Acta* **40**, 249–251 (1976).
- Eppley, R. W., Renger, E. H. & Betzer, P. R. *Deep-Sea Res.* **30**, 311–323 (1983).
- Saino, T. & Hattori, A. *Nature* **283**, 752–754 (1980).
- Menzel, D. W. & Ryther, J. H. *Deep-Sea Res.* **7**, 282–288 (1961).
- Honjo, S. in *Polar Oceanography, Part B: Biology, Chemistry and Geology* (ed. Smith, W. O. Jr) 687–739 (Academic, New York, 1990).
- Minagawa, M. & Wada, E. *Geochim. Cosmochim. Acta* **48**, 1135–1140 (1984).
- Wada, E., Terazaki, M., Kabaya, Y. & Nemoto, T. *Deep-Sea Res.* **34**, 829–841 (1987).
- Fry, B. *Limnol. Oceanogr.* **33**, 1182–1190 (1988).
- Deuser, W. G. *et al. Deep-Sea Res.* **37**, 1331–1343 (1990).

19. Saino, T. & Hattori, A. *Deep-Sea Res.* **34**, 807–827 (1987).
20. Gaebler, O. H., Vitti, T. G. & Vukmirovich, R. *Can. J. Biochem.* **44**, 1249–1257 (1966).
21. Bacon, M. P., Huh, C., Fleer, A. P. & Deuser, W. G. *Deep-Sea Res.* **32**, 273–286 (1985).
22. Siegel, D. A., Granata, T. C., Michaels, A. F. & Dickey, T. D. *J. geophys. Res.* **95**, 5305–5311 (1990).
23. Macko, S. A., Fogel, M. L., Hare, P. E. & Hoering, T. C. *Chem. Geol.* **65**, 79–92 (1987).
24. Lee, C. & Cronin, C. *J. mar. Res.* **42**, 1075–1097 (1984).
25. Ittekkot, V., Deuser, W. G. & Degens, E. T. *Deep-Sea Res.* **31**, 1057–1069 (1984).
26. Druffel, E. R. M. & Williams, P. M. *Nature* **347**, 172–174 (1990).
27. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep-Sea Res.* **34**, 267–285 (1987).
28. Bratbak, G. *Appl. envir. Microbiol.* **49**, 1488–1493 (1985).
29. Lee, S. & Fuhrman, J. A. *Appl. envir. Microbiol.* **53**, 1298–1303 (1987).
30. Liu, K.-K. & Kaplan, I. R. *Limnol. Oceanogr.* **34**, 820–830 (1989).
31. Altabet, M. A. *Limnol. Oceanogr.* **34**, 902–909 (1990).

ACKNOWLEDGEMENTS. We thank the crews of the RVs *Atlantis II*, *Meteor* and *Weatherbird* for field operations, E. Connors and L. Priest for technical assistance, and B. Fry for scientific consultation. The Ocean Sciences Division of the U.S. NSF funded this work through grants to M.A.A., W.G.D. and S.H.

Anhydrite-bearing pumices from Mount Pinatubo: further evidence for the existence of sulphur-rich silicic magmas

A. Bernard*, D. Demaiffe*, N. Mattielli* & R. S. Punongbayan†

* Department of Geology, CP 160, Université Libre de Bruxelles, 50 Avenue F. D. Roosevelt, 1050 Brussels, Belgium

† PHIVOLCS, Philippine Institute of Volcanology and Seismology, Hizon Bldg, 29 Quezon Avenue, Quezon City, Philippines

THE eruption of El Chichón in 1982 showed that relatively small but sulphur-rich eruptions from calc-alkaline volcanoes can produce long-lived stratospheric clouds of sulphate aerosols^{1,2}, which affect the global climate^{3,4}. Here we report the presence of primary anhydrite (CaSO₄) phenocrysts in dacitic pumice clasts from the 14–15 June eruption of Mount Pinatubo, which clearly shows that the Mount Pinatubo magma is also rich in sulphur. The post-eruptive sulphur content of the Pinatubo pumices ranges from 0.37 to 0.48 wt% SO₃. The considerable amount of sulphate aerosol injected into the stratosphere by the Mount Pinatubo eruptions⁵ should lead to a measurable cooling of the Earth's surface over the next few years, and could also trigger heterogeneous chemical reactions leading to stratospheric ozone depletion⁶. This new eruption of a sulphur-rich silicic magma thus shows that the El Chichón eruption is not unique, and that climate-modifying eruptions of this type may be more common than previously believed.

After 635 years of inactivity⁷, Mount Pinatubo (100 km NW of Manila in the Philippines) erupted violently on 14–15 June 1991. The climactic phase of the eruption lasted about 36 hours. The heights of the plinian column based on weather satellite data were estimated at 17–19 km on 12–13 June, at 20–25 km on 14 June and reached a maximum of 25–30 km on 15 June⁷. Heavy ash falls occurred near (0–40 km) the volcano and, carried by the strong winds of the typhoon Yunya, ash falls were observed as far away as Thailand and Singapore to the west. During the eruption, voluminous pyroclastic flows travelled down radiating stream channels, extending as far as 15–16 km from the main crater, and completely filling channels close to the volcano to depths of >100 m. According to data obtained with the TOMS instrument⁵, more than 19 million tons of SO₂ were injected at an altitude of 20–30 km by the June eruptions of Mount Pinatubo, which is nearly three times the amount of SO₂ outgassed by El Chichón volcano in 1982.

Preliminary estimates of the volume of volcanic debris show that 7 km³ of pyroclastic-flow deposits were erupted (G. M. Besana, personal communication). The volume of pyroclastic-fall material deposited on land is estimated at 0.6 km³. This is a minimum value as a large part of the ash fell in the sea. Although much larger than the recent eruptions of Mount St

Helens (1980) and El Chichón (1982), the Mount Pinatubo eruption, with an estimated volume erupted equivalent to 3–5 km³ of dense rock¹⁹, is not the largest of this century, falling behind Cerro Azul (1932), Katmai (1912) and Santa Maria (1902).

Thin sections were prepared from pumice clasts collected two weeks after the eruption. PIN-1 was collected from the top of a still-degassing pyroclastic-flow deposit in the Marella river (9 km SW of the volcano summit). PIN-2 and PIN-3 were collected from tephra-fall layers at more distant locations from the volcano. The three samples are homogeneous in whole-rock composition, and display porphyritic textures with large plagioclase and hornblende phenocrysts. All three samples contain anhydrite as phenocrysts and micro-phenocrysts. The anhydrite crystals (0.2–0.3 mm with one crystal up to 0.5 mm across) are scattered in the glassy groundmass. Most crystals are euhedral to subhedral with sharp contacts against the groundmass; no reaction coronae have been observed. Some crystals display lamellar twinning. Like the other phenocrysts, anhydrite contains abundant inclusions of apatite (needles and short prisms), Fe–Ti oxides, glass and, less commonly, tiny (<5 µm) zircons (observed in the scanning electron microscope). Anhydrite forms clusters with apatite (Fig. 1), as was noted in the El Chichón trachyandesite pumices, which could indicate that these two sulphur-bearing minerals nucleated upon one another⁸.

Euhedral plagioclase phenocrysts (3–5 mm, exceptionally up to 1 cm) typically have broken edges, and display growth zones with both normal and oscillatory zoning and complex twinning. Plagioclase composition varies widely from grain to grain; average composition (optical determination) is in the range An₃₅ to An₄₅. Some phenocrysts contain needles of apatite as well as glass inclusions with gas bubbles. Plagioclase microlites are uncommon. Many hornblende phenocrysts are euhedral, although others are broken. They contain many inclusions of apatite and Fe–Ti oxides and, less commonly, relicts of a deep-brown biotite. Titanium-poor magnetite and ilmenite occur together as microphenocrysts; they also contain apatite and zircon inclusions. Spinel, which was observed in the anhydrite-bearing El Chichón pumices, is completely lacking. Rare quartz is present as large rounded grains. Besides the apatite needles observed as inclusions in phenocrysts of plagioclase, hornblende and anhydrite, apatite also occurs as isolated euhedral prisms (0.1 mm) in the groundmass. Clinopyroxene (augite-ferro-augite) is rather uncommon, and only appears as small shattered crystals (<150 µm). Very rare pyrrhotite with detectable

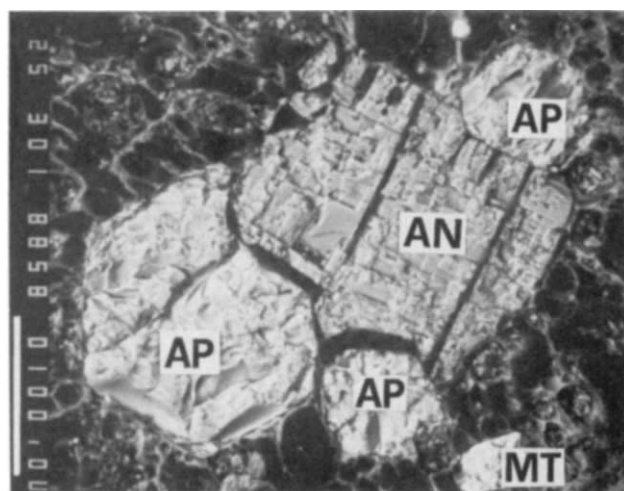


FIG. 1 Scanning electron micrograph (backscattered e⁻ mode) of the Pinatubo pumice showing a typical cluster of anhydrite (AN) and apatite (AP) crystals. MT is a Ti-poor magnetite. Scale bar, 100 µm.

TABLE 1 X-ray fluorescence analysis of whole-rock pumices

Weight %	PIN-1	PIN-2	PIN-3	Chichón ⁸	MSH ¹⁵
SiO ₂	64.08	64.35	64.24	55.88	63.03
TiO ₂	0.46	0.49	0.48	0.71	0.64
Al ₂ O ₃	15.71	15.37	15.59	17.80	17.51
Fe ₂ O ₃	2.15	2.11	2.30	2.59	1.82
FeO	2.01	2.07	1.97	3.42	2.66
MnO	0.11	0.11	0.11	0.19	0.08
MgO	2.59	2.37	2.51	2.18	2.10
CaO	5.28	5.15	5.17	7.91	5.20
Na ₂ O	4.67	4.80	4.44	4.46	4.59
K ₂ O	1.57	1.55	1.56	2.80	1.29
P ₂ O ₅	0.15	0.15	0.17	0.43	0.14
SO ₃	0.40	0.37	0.48	1.24	0.0075
Li	0.73	0.57	1.02	0.10	0.53
Total	99.91	99.46	100.04	99.71	99.68
Trace elements by X-ray fluorescence in p.p.m.					
Cl	473	476	523	1,288	400

FeO/Fe₂O₃ was determined by wet chemical analyses. Li, loss on ignition obtained by calcination of the sample at 950 °C after a preliminary drying of the powder in an oven at 110 °C for 2 h.

amounts (>1 wt%) of copper and nickel is present in the silicate phenocrysts.

The absence of reaction coronae, as well as the observation of frequent inclusions within anhydrite phenocrysts, strongly suggest that anhydrite was in equilibrium with the silicate melt at the time the magma erupted. Phase-equilibrium experiments in hydrous silicate systems have shown that anhydrite is a stable magmatic phase for oxygen fugacities about 1 log unit above the Ni-NiO reference buffer^{9,10}. These f_{O_2} conditions are typical of many calc-alkaline volcanic suites, especially those containing hornblende and biotite phenocrysts^{11,12}. The f_{O_2} of the Pinatubo magma estimated from the Fe₂O₃/FeO ratio¹³ of whole-rock pumices (Table 1) are around $10^{-10.5}$ for $T = 900$ °C, about 1.5 log unit above the Ni-NiO buffer.

But to crystallize anhydrite phenocrysts in detectable amounts, a magma must be strongly enriched in sulphur, with sulphur contents much higher than the values typically reported for calc-alkaline volcanic rocks (<150 p.p.m.)¹¹. Whole-rock chemical compositions of the Pinatubo pumices confirm this unusual sulphur enrichment (Table 1). Their SiO₂ (64 wt%) and K₂O (1.5 wt%) contents are typical for medium-K calc-alkaline dacites¹⁴. Although lower than in the El Chichón pumices, the post-eruptive sulphur contents of the Pinatubo samples are 50 times higher than for the Mount St Helens dacite (MSH)¹⁵ and clearly reflect an unusual enrichment. Considering the sulphur released to the stratosphere, the original pre-eruptive sulphur content was probably much higher. Estimations of the original sulphur content for the El Chichón magma are approximately double its residual value⁸: 2.6 wt% SO₃.

The occurrence of anhydrite as a primary crystallizing phase in a silicate magma has been rarely reported, mainly because of the relatively high solubility of this sulphate mineral in surface waters. As was observed for the El Chichón pumices, tropical rains can leach most of the primary anhydrite in less than one year⁸. But anhydrite is probably the only way to determine that ancient magmas were enriched in sulphur. Both matrix glasses and trapped glass inclusions in phenocrysts from the El Chichón pumices revealed very low sulphur contents⁸ (<0.1 wt%), probably because of the limited solubility of sulphur in silicate melts at upper-crustal T - P conditions^{8,10}.

Iron-rich basaltic effusive magmas are generally considered to be much more S-rich than their Fe-poor silicic and more explosive counterparts^{3,16-18}. The eruption of Mount Pinatubo involved a medium-K dacite with a more conventional calc-alkaline composition than the El Chichón high-K trachyandesite. This shows that the El Chichón event is not unique in

its high sulphur enrichment and that sulphur enrichment of calc-alkaline magma may be more common than previously believed. □

Received 2 August; accepted 1 October 1991.

- Pollack, J. B., Toon, O. B., Danielsen, E. F., Hofman, D. J. & Rosen, J. M. *Geophys. Res. Lett.* **10**, 989-992 (1983).
- Hofmann, D. J. & Rosen, J. M. *J. geophys. Res.* **92**, 9825-9830 (1987).
- Rampino, M. R. & Self, S. *Nature* **310**, 677-679 (1984).
- Self, S. & Rampino, M. R. *EOS* **69**, 83-86 (1988).
- Bluth, G., Doiron, S., Krueger, A., Schnetzler, C. & Walter, L. *Geophys. Res. Lett.* (in the press).
- Pitari, G., Visconti, G. & Rizi, V. *Geophys. Res. Lett.* **18**, 833-836 (1991).
- Global Volcanism Network Bull.* **16**, 2-7 (1991).
- Luhr, J. F., Carmichael, I. S. E. & Varekamp, J. C. *J. Volcan. geotherm. Res.* **23**, 69-108 (1984).
- Carroll, M. & Rutherford, M. J. *J. geophys. Res.* **90c**, 601-612 (1985).
- Luhr, J. F. *J. Petrology* **31**, 1071-1114 (1990).
- Gill, J. B. *Orogenic Andesites and Plate Tectonics* (Springer, Berlin, 1981).
- Carmichael, I. S. E. *Contr. Miner. Petrol.* **14**, 39-64 (1967).
- Kilinc, A., Carmichael, I. S. E., Rivers, M. L. & Sack, R. O. *Contr. Miner. Petrol.* **83**, 136-140 (1983).
- Basaltic Volcanism Study Project 193* (Pergamon, New York, 1981).
- Rose, W. I. Jr, Wunderman, R. L., Hoffman, M. F. & Gale, L. J. *J. Volcan. geotherm. Res.* **17**, 133-157 (1983).
- Sigurdsson, H. *EOS* **63**, 601-602 (1982).
- Haughton, D. R., Roeder, P. L. & Skinner, B. J. *Econ. Geol.* **69**, 451-467 (1974).
- Devine, J. D., Sigurdsson, H., Davis, A. N. & Self, S. *J. geophys. Res.* **89**, 6309-6325 (1984).
- Scott, W. E., Hoblitt, R. P., Daligdig, J. A., Besana, G. & Tubianosa, B. S. *EOS abstr.* (in the press).

ACKNOWLEDGEMENTS. The manuscript was improved by the review of J. F. Luhr. We thank U. Callebaut and S. Wauthier from the Belgium Embassy in Manila for their advice and support; J. C. Duchesne and G. Bologne from the Collectif Interuniversitaire de Géochimie Instrumentale for the XRF analyses and C. Chaval for the wet chemical analyses. The field work in the Philippines was supported by the Belgian NSF (FNRS).

Geochemical evidence for formation of the Bay of Islands ophiolite above a subduction zone

Don Elthon

Department of Chemistry, University of Houston, Houston, Texas 77204-5641 USA

THE Bay of Islands (BOI) complex in western Newfoundland is one of the world's best-known ophiolites, and has generally been interpreted as a fragment of ocean floor emplaced on the eastern margin of North America^{1,2}. The similarities of the BOI to lithosphere formed at mid-ocean ridges has led many investigators to use the BOI as an onland laboratory in which to study the geochemical, tectonic and geophysical properties of the oceanic lithosphere¹⁻⁹. There have been, nonetheless, a few suggestions that the BOI ophiolite formed above a subduction zone¹⁰⁻¹², as have most other ophiolites¹³⁻¹⁵. Here I report the existence of a substantial tantalum depletion in BOI magmas; Jenner *et al.*¹⁶ demonstrate that BOI magmas have a niobium depletion as well. These depletions are characteristic of magmas erupted above subduction zones but not at mid-ocean ridges, indicating that the BOI is not, after all, a good analogue for lithosphere formed at mid-ocean ridges.

Ophiolites are a particular association of ultramafic rocks, gabbroic rocks, dolerite dykes, pillow lavas and sediments¹⁷. There are a few hundred ophiolites or ophiolite fragments known, but studies of three major ophiolites (the Semail ophiolite in Oman, the Troodos ophiolite in Cyprus and the BOI) have dominated research in this field. Although early studies of ophiolites suggested that they formed at mid-ocean ridges, recent studies^{13-15,18-22} have shown that most ophiolites formed in a supra-subduction zone (SSZ) environment—that is, the wedge-shaped region above a subduction zone where arc, back-arc and fore-arc magmatism occurs. This interpretation is based on the recent discoveries of lavas in ophiolites^{13,15,18-20} (and their related cumulates^{21,22}) that are unlike mid-ocean-ridge basalts (MORB), but are comparable to magmas found in SSZ environments.

TABLE 1 Whole-rock analyses of dykes from North Arm Mountain, Bay of Islands ophiolite

Rock type	Dol	Dol	Dol	Dol	Q-Di	Q-Di	Plgr	Plgr	Normalizing values
Major elements (wt. %)									
SiO ₂	50.37	52.94	50.82	50.80	57.16	59.47	76.99	78.19	
TiO ₂	0.90	1.00	1.56	1.81	1.54	1.26	0.13	0.07	
Al ₂ O ₃	16.89	16.97	15.15	14.63	16.28	14.77	13.07	12.81	
Fe ₂ O ₃	9.23	9.58	11.04	13.03	11.63	11.27	1.12	1.00	
MnO	0.15	0.13	0.14	0.27	0.22	0.22	0.01	0.01	
MgO	9.02	6.40	7.08	6.83	2.28	2.20	0.39	0.10	
CaO	10.76	9.35	10.81	8.15	5.20	4.59	1.99	1.16	
Na ₂ O	2.98	4.20	2.57	4.64	5.83	5.46	5.53	6.32	
K ₂ O	0.39	0.17	0.00	0.15	0.33	0.45	0.12	0.04	
P ₂ O ₅	0.06	0.08	0.15	0.18	0.36	0.36	0.00	0.01	
Total	100.75	100.82	99.32	100.49	100.83	100.05	99.35	99.71	
Trace elements (p.p.m.)									
Th	BDL	0.21	0.32	0.88	1.17	1.40	4.67	6.73	0.029
Ta	0.04	0.055	0.12	0.16	0.35	0.27	0.90	1.03	0.014
La	1.55	1.91	3.87	5.02	10.8	8.70	21.0	21.2	0.235
Ce	3.94	4.93	11.4	12.4	28.6	22.5	47.3	60.9	0.603
Sr	223	274	125	139	175	154	82	49	7.80
P	264	352	660	792	1,584	1,584	BDL	44	46
Hf	1.67	1.17	2.86	3.61	7.21	5.50	7.43	7.94	0.104
Zr	42	31	97	127	186	125	182	142	3.94
Sm	2.11	1.78	3.72	4.64	7.13	6.08	6.79	12.3	0.147
Eu	0.76	0.73	1.35	1.49	2.29	1.83	1.12	0.45	0.056
Ti	5,400	6,000	9,360	10,860	9,240	7,560	780	420	0.436
Tb	0.55	0.43	0.86	1.05	1.45	1.31	1.24	2.78	0.036
Y	24	20	34	44	65	59	62	137	1.56
Yb	1.95	1.85	3.67	3.91	6.18	5.71	6.89	13.55	0.163
Lu	0.32	0.28	0.55	0.60	0.92	0.86	1.04	1.92	0.024
Ni	107	40	35	25	4	6	16	5	
Cr	284	110	86	46	20	19	43	33	
Sc	33.9	34.7	41.2	39.4	22.4	25.3	2.06	2.93	
Co	35.8	35.0	37.3	39.5	16.7	15.9	4.4	0.92	
Ba	27	21	<27	64	63	61	92	79	

Dol=Dolerite (<55% SiO₂). Q-Di=Quartz diorite (55–65% SiO₂). Plgr=Plagiogranite (>65% SiO₂). BDL=Below detection limits. Some of the major element data are from refs 2 and 35.

The presence of chromium-rich spinels ($\text{Cr}/(\text{Cr} + \text{Al}) > 0.6$) in some ultramafic rocks of the BOI has led to the suggestion that this ophiolite might also have SSZ affinities^{10,11}. Some interpretations of the regional geology and chronology of western Newfoundland have further suggested that the BOI formed as the floor of a back-arc basin¹². Most investigators, however, have concluded that the BOI formed at a mid-ocean-ridge spreading centre^{1–9,15,19}, because BOI dolerite dykes and pillow lavas were indistinguishable from MORB^{1–3}, and BOI gabbroic rocks are comparable to cumulates sampled at mid-ocean ridges^{4,5,23}.

The trace element compositions of representative fine-grained dolerite, quartz diorite and plagiogranite dykes are reported in Table 1. It is important to note that all of these samples have been metamorphosed to some extent and that many elements are mobilized during this process to the extent that they are not reliable geochemical indicators. The data for relatively immobile trace elements, normalized to CI chondrite abundances²⁴, are shown in Fig. 1a, b and c. In the following discussion, reference is made to negative elemental anomalies, which are defined as depletions of the specified element relative to the adjacent elements in chondrite-normalized diagrams (Fig. 1a, b or c). The presence of negative anomalies for tantalum, niobium and/or titanium is indicative of SSZ magmatism^{15,25–29}.

The dolerite dykes have nearly flat patterns with a substantial negative Ta anomaly (Fig. 1a). Negative Ta anomalies also occur in the quartz diorites (Fig. 1b) and plagiogranites (Fig. 1c). The quartz diorites have an additional negative Ti anomaly; because this Ti anomaly begins to develop only in those rocks where Fe–Ti oxides crystallize and is never found in the less differentiated dolerites (see Fig. 1a), it is clear that the Ti anomaly is related to the onset of Fe–Ti oxide crystallization

rather than being intrinsic to the parental magma. The plagiogranites have additional negative anomalies for phosphorus, europium and titanium due to substantial crystallization of apatite, plagioclase and Fe–Ti oxides, respectively. It is important to note that the Ta depletion is found in all BOI dykes analysed and that the size of the Ta anomaly remains almost uniform (chondrite-normalized $\text{Ta}/((\text{Th} + \text{La})/2) = 0.45 \pm 0.1$) throughout the dolerite/quartz diorite/plagiogranite series of rocks and is not modified by the >95% crystallization inherent in this sequence of rocks. The negative Ta anomaly is clearly intrinsic to the BOI parental magmas and is not an artefact of crystallization processes; the negative Ta anomaly apparently reflects an earlier depletion of the BOI mantle in Ta, rather than the presence of exotic residual phases (such as rutile, perovskite and sphene^{25–29}) that preferentially retain Ta, because these exotic phases have not been observed in my examination of residual BOI mantle samples. It is important to note that Jenner *et al.*¹⁶ demonstrate small negative Nb anomalies in BOI dolerites and much larger negative Nb anomalies in trondhjemites. Data presented in Fig. 1a–c and by Jenner *et al.* clearly show that BOI magmas are characterized by substantial negative Ta and Nb anomalies, comparable to those typically found within SSZ magmas^{25–32}.

Further evidence that the BOI has an SSZ signature rather than a MORB origin is the depletion of Ta compared with Hf and Th²⁷ (Fig. 2). The BOI samples plot within the SSZ field (D) rather than in the MORB fields (A or B). The La/Ta of BOI dolerites, furthermore, are typically 30–40 (Table 1) compared with 10–20 in MORBs²⁵.

It now appears that even the BOI, long regarded by most investigators as a mid-ocean ridge ophiolite, has a substantial SSZ component; this conclusion was not reached in earlier

studies¹⁻³ because Ta and Nb contents were not measured. The SSZ character of the BOI, however, is much more subtle than that found in most other ophiolites. There are, for example, no boninites or calc-alkaline magmas reported from the BOI, even though they are common in many other ophiolites^{13-15,18-20}. There is almost no orthopyroxene as a cumulus mineral in the BOI⁴⁻⁷, whereas it is abundant in many ophiolites with a significant SSZ component^{15,21,22}. The Cr/(Cr + Al) for spinels from most (>95%) BOI samples studied is <0.60, similar to spinels in MORBs and abyssal peridotites¹⁰; high-Cr spinels (Cr/(Cr + Al) > 0.60) that dominate many SSZ ophiolites are rare in the BOI. Other than the Ta anomaly described here and the Nb anomaly documented by Jenner *et al.*, other trace element characteristics of BOI dolerites are indistinguishable from MORB¹⁻³.

The Ta and Nb depletions do not negate the substantial geological evidence that the BOI formed by sea-floor spreading, but they do indicate that this spreading was in either a back-arc or fore-arc setting in a SSZ environment. The geochemical characteristics of the basaltic floor of many back-arc basins³⁰⁻³² are similar to those shown here (Figs 1a-c and 2) and in Jenner *et al.*¹⁶, with substantial depletions in Ta and Nb. The geochemical characteristics of lithosphere formed in fore-arc regions are not as well known, but seem similar to back-arc regions^{30,31,33} and would also be a suitable site for BOI formation.

In light of these new geochemical data, I suggest that reconstructions of the tectonic setting of the BOI during its formation need to show it in an SSZ environment. This re-interpretation of the BOI formation does not negate the proposal that the

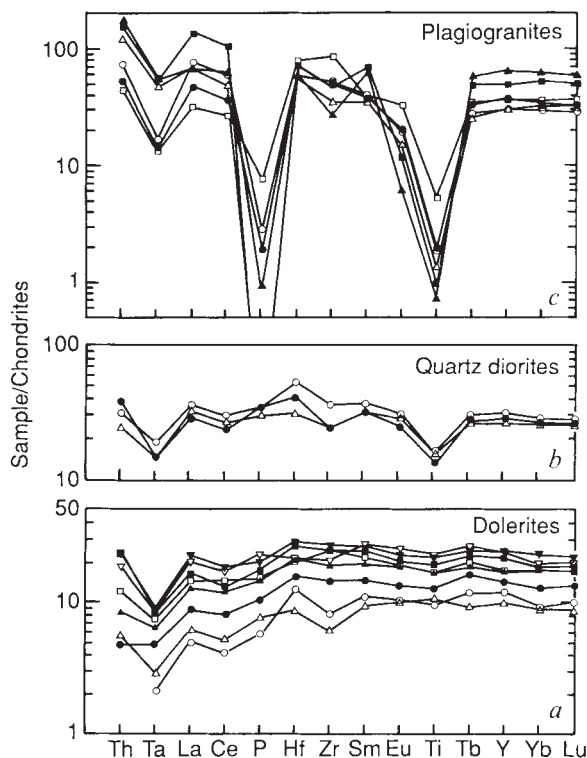


FIG. 1 Chondrite-normalized trace element diagram for dolerites, quartz diorites and plagiogranites from the North Arm Mountain massif of the Bay of Islands ophiolite. Representative analyses are reported in Table 1. The trace element data for these rocks are divided by the normalizing values in Table 1. Note the prominent negative Ta anomaly in all rocks. Data obtained by instrumental neutron activation analysis at the Johnson Space Center, except for Sr, P, Zr, Ti, Ni and Y which were obtained by X-ray fluorescence at the University of Houston (D.E., unpublished data, and ref. 35).

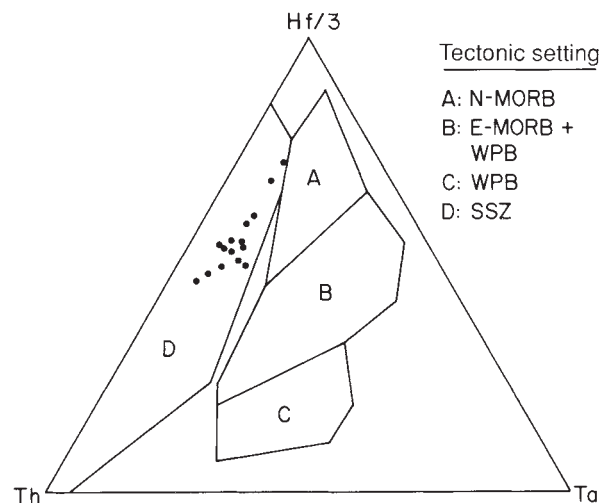


FIG. 2 Ta-Th-Hf discrimination diagram for rocks of different tectonic environments²⁷. Note that the Bay of Islands samples fall within the supra-subduction zone (SSZ) field, but not within the mid-ocean ridge basalt (MORB) or within plate basalt (WPB) fields. Back-arc and fore-arc basalts also plot within the SSZ field^{15,30}.

Coastal Complex and BOI ophiolites are related by fracture-zone tectonics and magmatism³⁴, but it does shift the location to a SSZ site. Earlier studies of the geophysical properties of the BOI^{8,9} also remain valid, but are possibly in need of slight revision to reflect potential differences in back-arc or fore-arc lithosphere formation compared with mid-ocean ridges. The role of high-pressure crystallization in the formation of BOI and oceanic lithosphere^{6,7}, however, was based on the belief that the liquids from which these minerals crystallized were MORB liquids. This belief is no longer valid; I now consider the role of high-pressure crystallization in the formation of the BOI^{6,7} to be inconsequential. The importance of high-pressure crystallization in the evolution of MORBs is not diminished by these results, but use of the BOI as a model for understanding these processes is no longer appropriate.

The BOI now joins the long list of other ophiolites with clear SSZ characteristics (see ref. 15). All the principal ophiolites (BOI, Troodos, Semail, Vourinos, Papua-New Guinea, Zambales and others) seem to have formed in SSZ environments. There is a need to re-examine those ophiolites that are believed to have MORB characteristics (such as Pindos, Liguria, Inzecca, Xigaze, Othris and Macquarie Island) to determine whether Ta and Nb depletions are characteristic for these ophiolites as well. With the exception of Macquarie Island, which is an uplifted ridge segment and is therefore likely to have a MORB character, the other ophiolites may well have a SSZ signature which was overlooked in earlier studies because it was sufficiently subtle that it could only be detected by the presence of negative Ta and Nb anomalies, such as noted here for the BOI. I suggest that ophiolites should generally be considered as having been formed in SSZ environments until it can be shown conclusively that Ta or Nb anomalies are not characteristic of the individual ophiolite. □

Received 1 August; accepted 17 October 1991.

- Malpas, J. *Phil. Trans. R. Soc. Lond. A* **288**, 527-546 (1978).
- Casey, J. F., Elthon, D. L., Siroky, F. X., Karson, J. A. & Sullivan, J. *Tectonophysics* **116**, 1-40 (1985).
- Suen, C. J., Frey, F. A. & Malpas, J. *Earth planet. Sci. Lett.* **45**, 337-348 (1979).
- Komarov, S. C., Elthon, D. & Casey, J. F. *Contrib. Mineral. Petrol.* **95**, 278-300 (1987).
- Komarov, S. C. & Elthon, D. *J. Petrol.* **31**, 1-50 (1990).
- Elthon, D., Casey, J. F. & Komarov, S. C. *J. geophys. Res.* **87**, 8717-8734 (1982).
- Elthon, D., Casey, J. F. & Komarov, S. C. *Geol. Soc. Lond. Spec. Publ.* **13**, 83-97 (1984).
- Collins, J. A., Brocher, T. M. & Karson, J. A. *J. geophys. Res.* **91**, 12520-12538 (1986).
- Karson, J. A. & Elthon, D. *Geology* **15**, 127-131 (1987).
- Dick, H. J. B. & Bullen, T. *Contrib. Mineral. Petrol.* **34**, 54-76 (1984).

11. Bedard, J. H. *J. Geol.* **99**, 225–249 (1991).
12. Searle, M. P. & Stevens, R. K. in *Ophiolites and Oceanic Lithosphere* (eds Gass, I.G., Lippard, S. J. & Shelton, A. W.) 303–319 (Blackwell, London, 1984).
13. Alabaster, T., Pearce, J. A. & Malpas, J. G. *Contrib. Mineral. Petrol.* **81**, 168–183 (1982).
14. Robinson, P. T., Melson, W. G., O'Hearn, T. & Schmincke, H.-U. *Geology* **11**, 400–404 (1983).
15. Pearce, J. A., Lippard, S. J. & Roberts, S. *Geol. Soc. Lond. Spec. Publ.* **16**, 77–94 (1984).
16. Jenner, G. A., Dunning, G. R., Malpas, J., Brown, M. & Brace, T. *Can. J. Earth Sci.* (in the press).
17. Penrose Field Conference Participants *Geotimes* **17**, 24–25 (1972).
18. Cameron, W. E. *Contrib. Mineral. Petrol.* **89**, 1–17 (1985).
19. Coish, R. A. in *Boninites* (ed. Crawford, A. J.) 264–287 (Unwin Hyman, London, 1989).
20. Menzies, M., Blanchard, D. P. & Xenophontos, C. *Am. J. Sci.* **280**, 329–344 (1980).
21. Cyprus Crustal Study Project *Geol. Surv. Canada Paper* 88–9 (1989).
22. Hawkins, J. W. & Evans, C. A. *Am. Geophys. Un. Monogr. Ser.* **27**, 95–123 (1982).
23. Beard, J. S. *Geology* **14**, 848–851 (1986).
24. Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–214 (1989).
25. Hole, M. J., Saunders, A. D., Marriner, G. F. & Tarney, J. *J. geol. Soc. Lond.* **141**, 453–472 (1984).
26. Pearce, J. A. in *Andesites: Orogenic Andesites and Related Rocks* (ed. Thorpe, R. S.) 525–598 (Wiley, Chichester, 1982).
27. Wood, D. A. *Earth planet. Sci. Lett.* **50**, 11–30 (1980).
28. Sun, S. S. & McDonough, W. F. *Geol. Soc. Lond. Spec. Publ.* **42**, 313–345 (1989).
29. Ryerson, F. J. & Watson, E. B. *Earth planet. Sci. Lett.* **86**, 225–239 (1987).
30. Saunders, A. D. & Tarney, J. *Geol. Soc. Lond. Spec. Publ.* **16**, 59–76 (1984).
31. Wood, D. A. *et al. Init. Rep. Deep Sea Drill. Proj.* **60**, 611–645 (1981).
32. Saunders, A. D. & Tarney, J. *Geochim. cosmochim. Acta* **43**, 555–572 (1979).
33. Ogawa, Y. & Naka, J. *Geol. Soc. Lond. Spec. Publ.* **13**, 291–301 (1984).
34. Karson, J. A. & Dewey, J. F. *Bull. Geol. Soc. Am.* **89**, 1037–1049 (1978).
35. Carter, K. E. thesis, Univ. of Houston, 150 (1985).

ACKNOWLEDGEMENTS. The INAA data were obtained at the Johnson Space Center through the courtesy of M. Lindstrom, D. Mittlefehldt and R. Martinez. This work was supported by the Texas Advanced Research Program.

Spinel elasticity and seismic structure of the transition zone of the mantle

Sally M. Rigden*, Gabriel D. Gwanmesia†, John D. Fitz Gerald*, Ian Jackson* & Robert C. Liebermann†

* Research School of Earth Sciences, Australian National University, GPO Box 4, Canberra, Australian Capital Territory, 2601, Australia

† Center for High Pressure Research and Department of Earth and Space Sciences, State University of New York, Stony Brook, New York 11794, USA

THERE is continuing debate about whether the upper mantle is chemically stratified, and whether the seismic discontinuity at 400 km depth represents a chemical boundary, a phase change of the $(\text{Mg, Fe})_2\text{SiO}_4$ component from the α phase (olivine) to the β phase, or a combination of both. Recent developments in high-pressure synthesis^{1,2} and ultrasonic interferometry^{3,4} have made possible measurements of elastic-wave velocities in small, polycrystalline samples of high-pressure phases. By combining our new acoustic measurements on the spinel (γ) phase of Mg_2SiO_4 with existing data for the α and β phases, we present here velocity profiles for the M_2SiO_4 component of the mantle (where M represents Mg or Fe) to depths of about 600 km. We find it to be unlikely that any seismologically observable velocity discontinuity at about 520 km depth can be attributed to this component, although the contrast in impedance (the product of density and velocity) might be sufficient for the $\beta \rightarrow \gamma$ transformation to be observed in long-period seismic reflection studies at near-normal incidence⁵. The velocity gradients in the transition zone, particularly for shear waves, are steeper than would be expected for simple adiabatic compression of likely mantle compositions, suggesting that alternative explanations including chemical heterogeneity and anelastic relaxation need to be explored.

Phase transformations in the $(\text{Mg, Fe})_2\text{SiO}_4$ system are commonly invoked to explain seismic discontinuities in the Earth's mantle, and recent seismic studies (for example, ref. 5) have focused attention on the possible existence of a velocity discontinuity at a depth of ~ 500 –550 km, a discontinuity often attributed^{5–7} to the $\beta \rightarrow \gamma$ transformation in $(\text{Mg, Fe})_2\text{SiO}_4$.

Single-crystal elasticity data^{6,7} have shown, however, that the velocity contrast between the β and γ phases is fairly small, only 1%, at ambient conditions. Values for the high-pressure and high-temperature elastic properties of these and other high-pressure phases that would allow calculation of velocity–depth profiles for the deep Earth have until now been unavailable and have therefore been inferred indirectly^{8–10}.

Recently, dense isotropic polycrystals of the γ phase of Mg_2SiO_4 have been hot-pressed at a pressure $P = 19.5$ GPa in a 2,000-ton uniaxial split sphere apparatus (USSA-2000) using techniques described elsewhere^{1,2}. Several specimens recovered at ambient conditions have bulk densities within 0.3% of the X-ray density of $\gamma\text{-Mg}_2\text{SiO}_4$ (3.559 g cm^{-3}). Light microscopy of a petrographic section of one such specimen (no. 1,039) indicates that it consists of more than 95% of the optically isotropic γ phase. Further investigation by transmission electron microscopy revealed that these spinel regions are made up of equant, well-formed grains of variable size up to $0.5 \mu\text{m}$ (Fig. 1). Grain boundaries and triple points are essentially free of pores and cracks. A small volume of optically anisotropic material consists of minor amounts of the α (olivine) and β phases ($<4\%$ and $<1\%$, respectively). Assuming the texture of this specimen to be representative, we conclude that specimens prepared in this manner are well sintered. The absence of cracks is confirmed by the close agreement between acoustic velocities measured at ambient pressure and those extrapolated from the high pressure experiments described below.

Another cylindrical specimen (no. 1,044) was selected for acoustic measurements. Two-way travel times were measured at pressures up to 3 GPa at ambient temperature for both shear and compressional waves by ultrasonic interferometry in the frequency range ~ 20 to ~ 70 MHz. The buffer-rod technique for use with jacketed polycrystalline specimens and the conversion of travel times to velocities has been described previously^{3,4,10–12}. Velocity–pressure data for $\gamma\text{-Mg}_2\text{SiO}_4$ are presented in Fig. 2. Below ~ 1.5 GPa, the results are mildly contaminated by the effects of imperfect mechanical coupling across a thin gold bond. Above $P = 1.5$ GPa, the velocity varies linearly with pressure.

To facilitate consistent extrapolation of these results to the much higher pressures of the transition zone, the shear and compressional moduli have been fitted to linear combinations

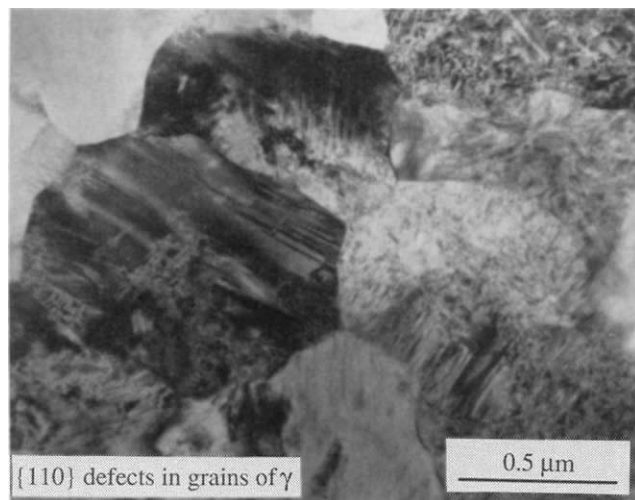


FIG. 1 Well-formed grains of Mg_2SiO_4 spinel seen by transmission electron microscopy, showing the excellent sintering of Mg_2SiO_4 spinel polycrystal (bright field, ion-thinned specimen from run 1,039). Note the absence of cracks or voids and pores on grain boundaries. The complex planar defects imaged inside each grain have contrast properties consistent with an origin as cation stacking faults²⁶ on $\{110\}$ spinel planes.

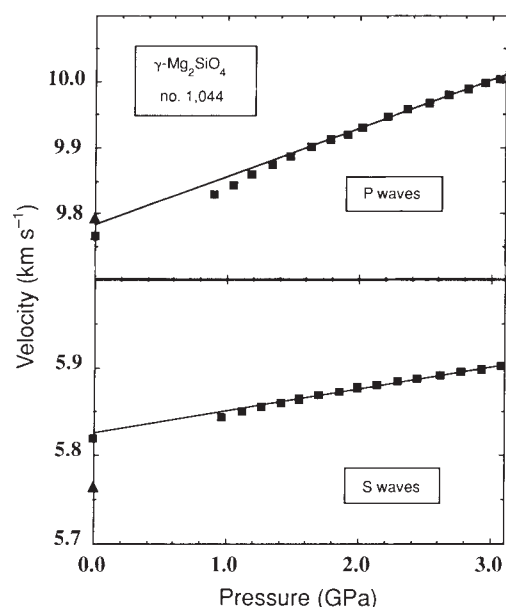


FIG. 2 Measured compressional (P) and shear (S) wave velocities (shaded squares) for polycrystalline Mg_2SiO_4 spinel specimen no. 1,044 as a function of pressure up to 3 GPa and at room temperature. Linear trends of the high-pressure data above 1.5 GPa are in good agreement with measured values at 1 atm and with the Hashin-Shtrikman aggregate velocities ($\blacktriangle$) calculated from single-crystal Brillouin scattering data⁷.

of functions of the eulerian strain. The coefficients of these functions are related to the adiabatic bulk modulus (K) and shear modulus (G) and their pressure derivatives¹³. Values obtained for the γ -phase specimen are $(\partial K/\partial P)_T = 5.0$ and $(\partial G/\partial P)_T = 1.7$. These pressure derivatives of elastic moduli are very similar to those of the β^{11} and α phases^{14,15}.

Calculation of the pressure dependence of P- and S-wave velocities for an M_2SiO_4 component to pressures equivalent to ~ 600 km depth in the Earth indicates that the effect of pressure alone is to decrease slightly (to below 1%) the small velocity contrast that exists between the β and γ phases at ambient laboratory conditions. Quantification of the contribution of the $\beta \rightarrow \gamma$ phase transition to any velocity discontinuity in the mantle is hampered because the temperature dependence of elasticity has not yet been measured for these phases. Some bounds may, however, be placed on the velocity and density contrast by investigating a range of plausible temperature derivatives. Our confirmation of the prediction⁹ that the pressure derivatives of elastic moduli for both β and γ phases were olivine-like strengthens the suggestion⁹ that the temperature derivatives might also be olivine-like and not very different from each other. If this were the case, however, comparison of the velocity discontinuities generated at the $\alpha \rightarrow \beta$ phase transition with the seismologically observed discontinuities would yield an inferred M_2SiO_4 content of 40–53% in the mantle at this depth on the basis of compressional velocities, but only 35% on the basis of shear velocities⁹. It was this apparent discrepancy that led to a different approach to estimating temperature derivatives¹¹. Pairs of $(\partial K/\partial T)_P$ and $(\partial G/\partial T)_P$ were calculated for the β phase that were consistent with the magnitudes of the seismologically observed velocity discontinuities for both P and S waves and for a range of M_2SiO_4 contents of the model mantle.

Calculation of the amplitude of the velocity discontinuity near 520 km depth from the $\beta \rightarrow \gamma$ transition was carried out using a range of plausible temperature derivatives^{8,9,11,14,15} in combination with the measured pressure derivatives. At 18 GPa, discontinuities for the M_2SiO_4 component are predicted to be: $\Delta V_P = 1\text{--}2\%$, $\Delta V_S = 0.8\text{--}1.5\%$ and $\Delta \rho = 2.5\text{--}3\%$ where ρ is density, and

V_P and V_S are the velocities of the P and S waves. The magnitudes of these discontinuities can be changed considerably only if the derivatives are substantially different ($>10\%$) for the β and γ phases. These values will be reduced by the dilution with non- M_2SiO_4 components amounting typically to 40–60%. Furthermore, recent experimental work^{16,17} on the system $(\text{Mg}, \text{Fe})_2\text{SiO}_4$ suggests that the transition occurs over a depth interval of ~ 50 km for the $(\text{Mg}_{0.9}, \text{Fe}_{0.1})_2\text{SiO}_4$ composition appropriate for the mantle. Although velocity discontinuities expected at this depth from the $\beta \rightarrow \gamma$ transition are small, the effect of the density contrast is to produce impedance contrasts of 2.5–5% for P waves and 3–4.5% for S waves. The inference of a discontinuity from near-vertical-incidence long-period records, for which there is no compelling evidence (see, for example, ref. 5) in short-period and refraction data (which are less sensitive to density contrast), can be reconciled with the dominant contribution of density to the impedance contrast. Thus the $\beta \rightarrow \gamma$ phase transition cannot be ruled out as the cause of the 520-km discontinuity, although the velocity discontinuity may be too small to be generally observable, particularly with short-period waves.

It is instructive to examine expected transition-zone velocity gradients for the M_2SiO_4 component for a range of plausible values of $(\partial K/\partial T)_P$ and $(\partial G/\partial T)_P$. Much earlier work has focused on the size of the velocity contrast at 400 km and its implications for the M_2SiO_4 content of the transition zone^{8,9,11} with recent estimates varying from 40% (refs 7, 9) to more than 70% (ref. 8). From our earlier calculations¹¹ and Fig. 3, it is clear that the size of the 400-km velocity discontinuity (and thus the inferred M_2SiO_4 content of a homogeneous mantle) is strongly influenced by the choice of temperature derivatives for the β -phase.

High-velocity gradients in the transition zone are commonly ascribed to continuous phase changes. The important reactions that are expected between 400 and 600 km are the completion of the progressive pyroxene to garnet reactions (although these are expected to be complete by 450 km depth¹⁸) and the $\beta \rightarrow \gamma$ phase transition between 490 and 540 km (ref. 17). The elastic properties of the silicate garnets at high pressure and temperature have not been measured yet; hence the discussion is confined primarily to the M_2SiO_4 component. Finite-strain adiabats are calculated using a similar approach to that in ref. 9 in correcting for the effect of temperature on densities and elastic moduli. The calculated profiles presented in Fig. 3 (solid

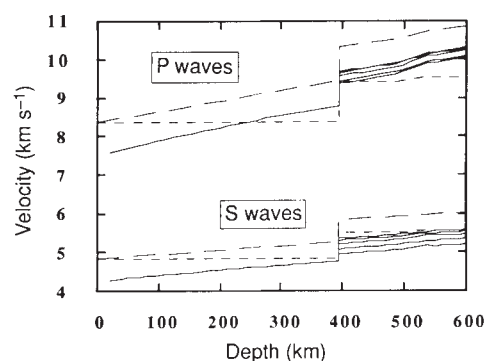


FIG. 3 Calculated acoustic profiles for $(\text{Mg}_{0.9}, \text{Fe}_{0.1})_2\text{SiO}_4$. Three different calculations are presented. The short-dashed line uses velocities measured at 1 atm. The long-dashed line shows the effect of correction for pressure using the measured pressure derivatives for olivine¹⁴, β -phase¹¹ and spinel (this work). The solid curves include correction for temperature along finite strain adiabats with a potential temperature of 1,400 °C for a range of estimated temperature derivatives of elastic moduli^{8,9,11,14,15}. The effect of changing temperature derivative is more pronounced on the magnitude of the velocity discontinuity near 400 km than on velocity gradients between 400 and 600 km.

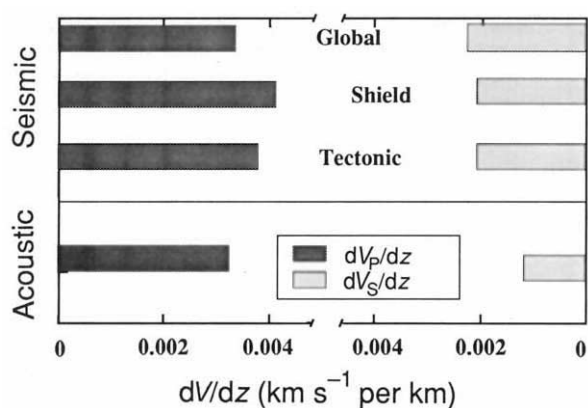


FIG. 4 Average velocity gradients for P and S waves through the transition zone between 400 and 600 km using seismic models using global travel-time data¹⁹ and synthetic body-wave studies for shield (S25²⁰, SNA²¹) and tectonic (GCA²², CJF²³, TNA²¹) and acoustic models from Fig. 3. Note that although P-wave seismic gradients are generally comparable with the acoustic gradients for the M_2SiO_4 component, S-wave seismic gradients are almost twice as great as the acoustic values.

curves) are based on our measured values of $(\partial K/\partial P)_T$ and $(\partial G/\partial P)_T$ for the β and γ phases, and various estimates^{8,9,11} of $(\partial K/\partial T)_P$ and $(\partial G/\partial T)_P$. The velocity gradients between 400 and 600 km are relatively insensitive to the values chosen for the temperature derivatives.

In Fig. 4, we compare the average velocity gradients for P and S waves through the transition zone (400–600 km) from recent seismological models, based on global travel-time data¹⁹, and synthetic body-wave studies for shield^{20,21} and tectonic^{21–23} regions, with those calculated from acoustic data for a model mantle of M_2SiO_4 composition (from Fig. 3). It is clear from this comparison that although adiabatic compression of the β and γ phases in the transition zone might be sufficient to explain the observed seismic gradients for P waves, this is not the case for shear waves where the seismic gradients are twice as great as the acoustically derived values.

In principle, this discrepancy could result from our neglect of $MgSiO_3$ -rich majorite garnet which is expected to be the only other abundant phase throughout the bulk of the transition zone. But the progressive pyroxene-to-garnet transformation, expected to be complete within 50 km of the 400-km discontinuity for compositions ranging from peridotite to eclogite¹⁸, cannot influence the gradient below this depth. Furthermore, similar velocity gradients to those calculated for the β and γ phases are inferred for this garnet phase from a combination of recent measurements at ambient conditions²⁴ with plausible pyrope-like pressure and temperature derivatives. In the face of these difficulties in matching the seismologically inferred velocity gradients with a model mantle of uniform chemical composition, it has been suggested that compositional heterogeneity is required (see, for example, ref. 9). Specifically, this would require progressive increase with increasing depth of the proportion of phases that have a relatively high ratio of V_S/V_P . Alternatively, it is possible that the shear-wave velocity gradient is steepened in the transition zone as a consequence of the gradual recovery of a shear modulus deficit associated with anelastic relaxation²⁵ concentrated at shallower depths than 400 km where the geotherm more closely approaches the mantle solidus. □

- Rigden, S. M., Jackson, I., Niesler, H., Ringwood, A. E. & Liebermann, R. C. *Geophys. Res. Lett.* **15**, 605–608 (1988).
- Niesler, H. & Jackson, I. *J. acoust. Soc. Am.* **86**, 1573–1585 (1989).
- Shearer, P. M. *Nature* **344**, 121–126 (1990).
- Sawamoto, H., Weidner, D. J., Sasaki, S. & Kumazawa, M. *Science* **224**, 749–751 (1984).
- Weidner, D. J., Sawamoto, H., Sasaki, S. & Kumazawa, M. *J. geophys. Res.* **89**, 7852–7860 (1984).
- Bina, C. & Wood, B. J. *J. geophys. Res.* **92**, 4853–4866 (1987).
- Duffy, T. S. & Anderson, D. L. *J. geophys. Res.* **94**, 1895–1912 (1989).
- Rigden, S. M. & Jackson, I. *J. geophys. Res.* **96**, 9999–10006 (1991).
- Gwanmesia, G. D., Rigden, S. M., Liebermann, R. C. & Jackson, I. *Science* **250**, 794–797 (1990).
- Rigden, S. M., Gwanmesia, G. D., Jackson, I. & Liebermann, R. C. in *High Pressure Research: Applications to Earth and Planetary Sciences* (eds Syono, Y. & Manghnani, M. H.) (TerraPub, Tokyo, in the press).
- Birch, F. *J. geophys. Res.* **57**, 227–285 (1952).
- Kumazawa, M. & Anderson, O. C. *J. geophys. Res.* **74**, 5961–5972 (1969).
- Graham, E. K. & Barsch, G. R. *J. geophys. Res.* **74**, 5949–5960 (1969).
- Katsura, T. & Ito, E. *J. geophys. Res.* **94**, 15663–15670 (1989).
- Akagi, M., Ito, E. & Navrotsky, A. *J. geophys. Res.* **94**, 15671–15685 (1989).
- Irfune, T. & Ringwood, A. E. in *High-Pressure Research in Mineral Physics* (eds Manghnani, M. H. & Syono, Y.) (TerraPub, Tokyo; AGU, Washington DC 1987).
- Kennett, B. L. N. *Geophys. Res. Lett.* **18**, 1115–1118 (1991).
- Le Fevre, L. V. & Helmberger, D. V. *J. geophys. Res.* **94**, 17729–17765 (1989).
- Grand, S. P. & Helmberger, D. V. *Geophys. J. R. astr. Soc.* **76**, 399–438 (1984).
- Walck, M. C. *Geophys. J. R. astr. Soc.* **76**, 697–723 (1984).
- Walck, M. C. *Geophys. J. R. astr. Soc.* **81**, 243–276 (1985).
- Yeganeh-Haeri, Weidner, D. J. & Ito, I. *Geophys. Res. Lett.* **17**, 2453–2456 (1990).
- Jackson, I., Paterson, M. S. & FitzGerald, J. G. *Geophys. J. Int.* (in the press).
- Vaughan, P. J. & Kohstedt, D. L. *Phys. Chem. Miner.* **7**, 241–245 (1981).

ACKNOWLEDGEMENTS. We thank D. Weidner, B. Kennett and P. Shearer for helpful comments on an earlier manuscript. This work was partially supported by the NSF (as work carried out at the NSF Science and Technology Center for High Pressure Research) and by the U.S.–Australia Cooperative Research Program with funding from the NSF and the Australian Department of Industry, Technology and Commerce.

Laser-fusion $^{40}\text{Ar}/^{39}\text{Ar}$ dating of Bed I, Olduvai Gorge, Tanzania

R. C. Walter*, P. C. Manega†, R. L. Hay‡, R. E. Drake* & G. H. Curtis*

* Geochronology Center, Institute of Human Origins, 2453 Ridge Road, Berkeley, California 94709, USA

† Department of Geology and INSTAAR, University of Colorado, Boulder, Colorado 80309, USA

‡ Department of Geology, University of Illinois, Urbana, Illinois 61801, USA

BED I of Olduvai Gorge has yielded some of the most important hominid fossils in the world, but precise age estimates have been elusive. Here we report new dating results for Bed I, based on $^{40}\text{Ar}/^{39}\text{Ar}$ analyses of single mineral grains using the laser-fusion technique. These dates reveal that the fossiliferous deposits of middle to upper Bed I, a period of significant biological and climatic change, comprise an extremely brief interval of time, from 1.80 Myr to 1.75 Myr. Dates for lower Bed I suggest that the base of the Olduvai subchron, of the geomagnetic polarity timescale, may be around 100,000 yr older than the currently accepted value.

Despite extensive dating studies at Olduvai Gorge over the past 30 years^{1–4}, its detailed chronology is largely unknown. Previous studies found that Bed I and Bed II contained tephra suitable for K–Ar dating³. Subsequent work, however, revealed that most tephra produced anomalously old dates due to detrital contamination⁵. In fact, only Tuff IB, an ignimbrite, produced consistently reliable dates of about 1.84 ± 0.03 Myr (corrected for revised ^{40}K decay constants⁶). This report presents new geochronological data for Bed I based on $^{40}\text{Ar}/^{39}\text{Ar}$ analyses using the single-crystal laser-fusion technique (SCLF), a method well suited for detecting and avoiding contamination^{7,8}.

Palaeomagnetic studies revealed a period of normal polarity in the lavas and tuffs of Bed I and lower Bed II^{9–11}. K–Ar dates for lower Bed I^{3,4} placed this interval of normal polarity within the palaeomagnetically reversed Matuyama chron. Based on these data, the Olduvai subchron, now internationally recognized, was established and dated to between 1.87 and 1.67 Myr¹².

The Naabi ignimbrite, below Bed I (Fig. 1), contains phenocrysts of sanidine and xenoliths of crystalline basement material. A conventional K–Ar date for the Naabi yielded an age of

Received 7 June; accepted 17 September 1991.

- Gwanmesia, G. D., Liebermann, R. C. & Guyot, F. *Geophys. Res. Lett.* **17**, 1331–1334 (1990).
- Gwanmesia, G. D. & Liebermann, R. C. in *High Pressure Research: Applications to Earth and Planetary Sciences* (eds Syono, Y. & Manghnani, M. H.) (TerraPub, Tokyo, in the press).

5.01 Myr⁴, but Hay⁵ suggests that this is too old because of contamination. Indeed, SCLF analyses of sanidine feldspars from the Naabi (Tables 1, 2) yield an age of 2.029 ± 0.005 Myr. Ten grains were measured, one of which yielded an age of 2.8 Myr, possibly representing a partially degassed xenocryst; this grain was excluded from the final age calculation (Table 2). The remaining grains possess similar Ca/K ratios (Table 2) and exhibit a normal age distribution (Fig. 2).

A coarse feldspar crystal tuff (CFCT) occurs in the Main Gorge west of the Fifth Fault. SCLF analyses on feldspars from different localities yield concordant ages of 2.00 ± 0.02 Myr and 2.03 ± 0.01 Myr, respectively, and a mean age of 2.02 ± 0.02 Myr (Table 1). This age is similar to the Naabi, but Ca/K ratios for the CFCT feldspars are two times higher, suggesting a chemically distinct and separate eruption. Two of the ten grains measured from OG-5 (Loc. 66b) yield Pan-African ages of 460 and 740 Myr (Table 2).

Tuff IA is reworked tuff exposed west of the Fifth Fault above the Naabi⁵. Of the ten grains analysed from Tuff IA, seven are detrital Pan-African feldspars incorporated in the tuff during redeposition (Table 2). Two of the remaining three grains are anorthoclase feldspars (Ca/K = 0.49) that yield concordant SCLF ages of 1.98 ± 0.02 Myr. The third grain is a sanidine

feldspar (Ca/K = 0.03) with an SCLF age of 2.07 ± 0.03 Myr; its composition and age suggests that it is a contaminant from the CFCT. The youngest age component for Tuff IA, 1.98 Myr, is stratigraphically appropriate. Additionally, higher Ca/K ratios distinguish Tuff IA feldspars from either the Naabi Ignimbrite or the CFCT.

The Bed I lava yields a bulk $^{40}\text{Ar}/^{39}\text{Ar}$ age of 1.87 ± 0.02 Myr, which is consistent with its magnetostratigraphy. As the Bed I lava is variably altered, and previous K-Ar ages ranged from 1.88 to 2.10 Myr⁵, we consider 1.87 Myr a minimum estimate of the age of the Bed I lavas.

SCLF results for Tuff IB yield two different ages. Feldspars from the air-fall deposit directly beneath the Tuff IB ignimbrite are dated to 1.86 ± 0.01 Myr, whereas feldspars from pumice clasts within the ignimbrite, and from an air-fall deposit above the ignimbrite, yield ages of 1.80 ± 0.01 Myr (Table 1). Argon isotope correlation data support the bimodal age results (Table 1). These data suggest that an air-fall event preceded the eruption of the Tuff IB ignimbrite by as much as 60,000 yr, and that the ignimbrite was followed closely in time by a co-ignimbrite air-fall eruption. Alternatively, the high initial $^{40}\text{Ar}/^{36}\text{Ar}$ ratio (Table 1) indicates that feldspars from the lower Tuff IB air-fall deposit may contain excess radiogenic ^{40}Ar .

TABLE 1 Summary of mean SCLF dates for Bed I

Unit	Ca/K*	SCLF results				Ar isotope correlation results				MSWD	n
		±	Age (Myr)	±	Age (Myr)	±	(⁴⁰ Ar/ ³⁶ Ar) _i †	±			
Tuff IF	0.280	0.07	1.748	0.011	1.770	0.021	281.8	14.1	1.5	36	
			1.749	0.007	1.759	0.011					
Tuff IE‡	0.360	0.01	1.789	0.053	1.790	0.020	314.0	7.0	2.2	15	
			1.797	0.026	1.75§	0.020	310.0	13.0	0.8	10	
					1.87	0.020	295.0	9.0	0.2	5	
Tuff ID	2.730	1.63	1.762	0.010	1.750	0.024	296.8	13.8	2.3	29	
			1.764	0.014	1.740	0.015					
Tuff IC	0.360	0.01	1.764	0.023	1.790	0.085	294.0	17.0	0.9	13	
			1.761	0.028	1.790	0.066					
Tuff IB (I and U)¶	0.110	0.05	1.803	0.013	1.802	0.015	279.3	54.3	2.8	41	
			1.798	0.004	1.806	0.008					
Tuff IB (L)¶	0.080	0.04	1.862	0.015	1.853	0.018	323.3	26.0	4.1	44	
			1.859	0.007	1.865	0.007					
Lava#	2.420	0.85	1.863	0.006	1.790	0.071	301.0	4.2	1.7	12	
			1.865	0.020	1.745	0.100					
Tuff IA	0.490	0.03	1.976	0.015	1.970	0.160	318.0	29.0	0.0	2	
CFCT	0.030	0.01	2.014	0.023	2.017	0.043	302.0	22.6	0.6	11	
			2.018	0.019	2.043	0.017					
Naabi	0.001	0.00	2.029	0.005	2.030	0.010	297.0	64.0	0.2	9	

Preferred mean ages are boxed and bolded. The first line in each data set (for example Tuff IF) represents a simple average and standard deviation of n measurements. The second line (bold) represents the weighted mean and weighted uncertainty (1σ)²³⁻²⁵ for n measurements. Errors for individual grain ages (see Table 2) represent 1σ analytical uncertainties, and incorporate errors in J ²⁴. Errors cited for weighted mean ages are calculated using the weighted uncertainty formula of Samson and Alexander²⁵, which propagates the error in J from the uncertainty in individual analyses. Statistical treatment of the Ar isotope correlation method is similar to that described by McDougall²⁰. MSWD (mean square of weighted deviates) is a reduced χ^2 variable. If the fitting function is a good approximation of the parent function, then $S^2 \sim \sigma^2$, and MSWD ~ 1 . If the fitting function is not appropriate, then MSWD $\gg 1$. Values less than 1 do not imply a superior fit, but rather are a consequence of the uncertainty in determining the estimated variance S^2 ; the observed values for MSWD will fluctuate ~ 1 from experiment to experiment²⁶.

* Ca/K is derived by multiplying the $^{37}\text{Ar}/^{39}\text{Ar}$ ratio by 1.96.

† Initial $^{40}\text{Ar}/^{36}\text{Ar}$ derived from the Ar isotope linear regression.

‡ Combined feldspar data from three pumice clasts.

§ Refers to Ar isotope correlation results for the younger population in Fig. 2.

|| Refers to Ar isotope correlation results for the older population in Fig. 2.

¶ Tuff IB is a key marker bed traceable for over 13 km westward from the Second Fault⁵. At the Second Fault, it is a massive, partially welded, 3–4-m thick ignimbrite (Tuff IB (I)). It is underlain by a 5–10-cm thick bedded pumice lapilli-fall deposit, that grades upward into a fine, crystal-rich ash (Tuff IB (L)). A 10–15-cm thick bedded pumice lapilli was found to overlie the ignimbrite at one locality (Tuff IB (U)).

Lavas were analysed using the 'mini-bulk' method²⁷, whereby roughly 1–3 mg of prepared basalt were fused using a defocused laser beam.

Tuffs IC and ID overlie several important hominid localities, but previous attempts to date them by the K-Ar method have produced discordant results⁴. Both tephra units yield similar SCLF ages of 1.76 ± 0.02 Myr. Feldspars from Tuff IC have significantly lower Ca/K ratios than feldspars in Tuff ID, which is consistent with Hay's⁵ petrographic observations.

Tuff IE is an unwelded ignimbrite occurring near the top of Bed I⁵. Anorthoclase feldspars were extracted from three pumice clasts, and five grains were analysed from each clast. SCLF analyses on grains within each clast are broadly discordant, yielding ages ranging from 1.84 to 1.73 Myr, for a mean of 1.80 ± 0.03 Myr. A histogram and probability density function of all 15 grains, however, reveals two age populations (Fig. 2b), one at 1.76 ± 0.02 Myr ($n = 10$) and one at 1.87 ± 0.02 Myr ($n = 5$). Similarly, an inverse Ar correlation analysis reveals two age populations at 1.75 ± 0.02 Myr and 1.87 ± 0.02 Myr (Table 1).

An age of 1.75 to 1.76 Myr for Tuff IE is stratigraphically appropriate, and we accept this as a reasonable estimate of the eruptive age of Tuff IE. It is, however, unclear why bimodal

feldspar ages occur in pumices in primary context. An inverse Ar isotope correlation plot, for example, shows that the five older grains have an initial $^{40}\text{Ar}/^{36}\text{Ar}$ value indistinguishable from the atmospheric value of 295.5, indicating that excess argon is not the cause of the old ages. Furthermore, Ca/K values are similar for both age groups, implying that the older grains are chemically similar, at least with respect to Ca/K, to feldspars of the younger age group. Bogaard *et al.*⁸ recognized a similar bimodal age population from feldspars with equivalent Ca/K ratios in pumices from Quaternary tephra in the East Eifel volcanic field of Germany. We surmise that the magma that produced Tuff IE picked up older volcanic feldspars from the vent during eruption, and that these grains only partially degassed before cooling below the temperature for the retention of argon.

Previous K-Ar determinations on Tuff IF ranged from 1.71 to 20.1 Myr⁴. A SCLF analysis was performed on Tuff IF feldspars from the same vial that produced a K-Ar age of 5.67 Myr⁴. Our results show that five of six grains yield a mean age of

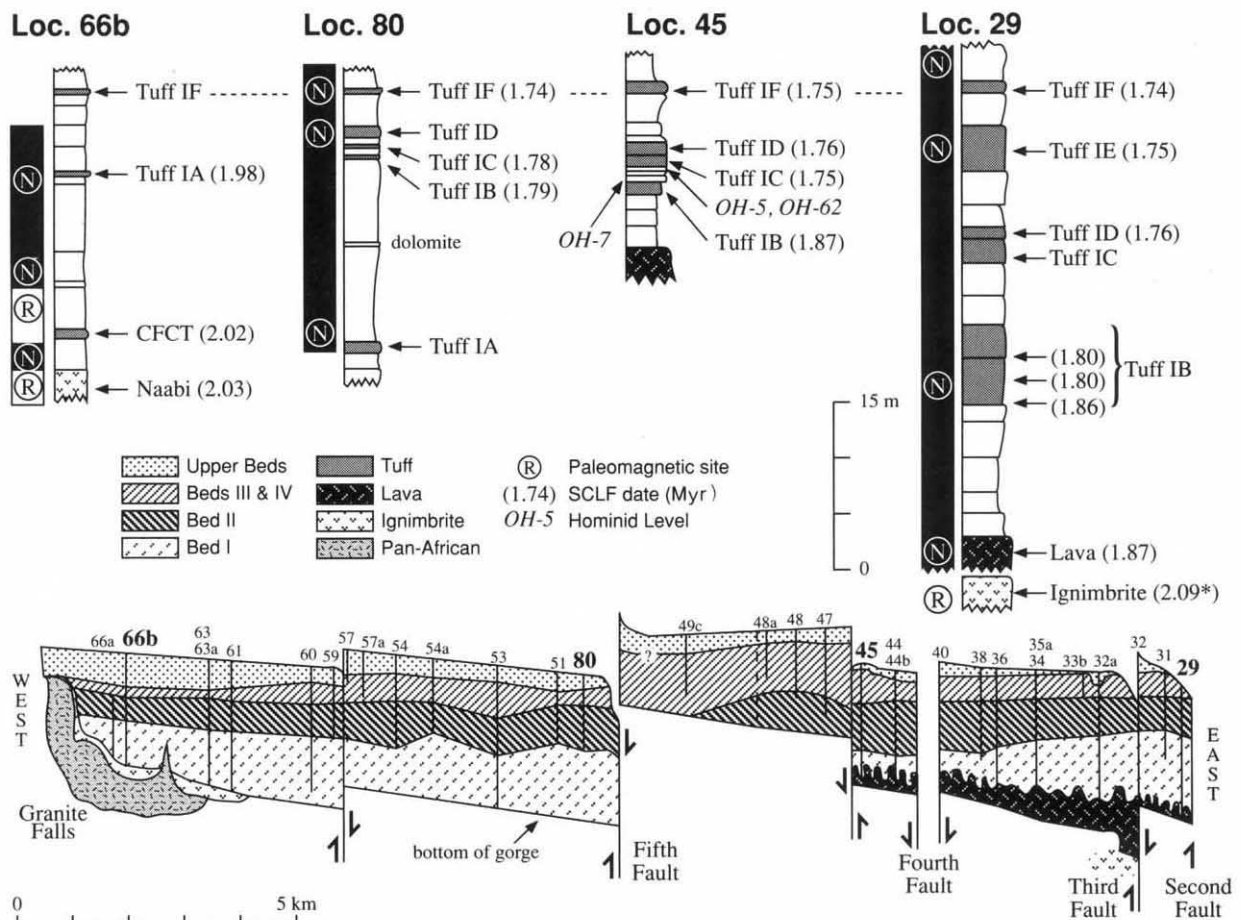


FIG. 1 A palinspastic reconstruction of Olduvai Gorge stratigraphy, with summary SCLF results for key Bed I sections. Olduvai Gorge is a 20-km long, narrow, deeply cut valley in the eastern Serengeti plain. It is composed of Pleistocene to Holocene sediments that overlie Plio-Pleistocene ignimbrites and lava flows. These volcanic deposits unconformably overlie metamorphic rocks of the Pan-African Mozambique belt. The entire depositional sequence is roughly 100 m thick, and contains lacustrine, lake margin, fluvial, and alluvial deposits interbedded with numerous primary and reworked tephra. The deposits are divided into seven beds, but stratigraphic correlations are locally complicated by faulting, abrupt facies changes and disconformities. Bed I deposits encompass five lithofacies (lavas and ignimbrites, lacustrine, lake margin, alluvial fan, and alluvial plain), and contain six major tephra units (Tuffs IA-IF) that enable stratigraphic correlations to be made throughout the gorge⁵. The eastern alluvial fan facies is composed of thick

volcanoclastic sediments and debris flows derived from the adjacent Ngorongoro volcanic highland. This clastic wedge thins to the west and interdigitates with fossiliferous lake margin sediments that characterize Bed I deposits around the junction of the Main and Side Gorges. Westward, lake margin sediments grade into fine grain lacustrine deposits. In the eastern alluvial fan facies (Loc. 29)⁵ Bed I—from Tuff IB to Tuff IF—is roughly 30 m thick. In the lacustrine deposits 10 km to the west (Loc. 80) the same stratigraphic sequence is roughly 6 m thick. The original stratigraphic sections and composite stratigraphies are from Hay^{5,28}. The locations of key hominid discoveries are shown in italics for Loc. 45. The reversely magnetized ignimbrite shown at the base of Loc. 29(*) actually occurs beneath the lava at the Third Fault; it was dated to 2.09 Myr by K-Ar and cited in ref. 4.

1.74 ± 0.03 Myr, whereas the sixth is 555 Myr. Additional SCLF analyses were performed on samples from the Second Fault (locality 5), from the Junction (localities 40 and 45c), and from west of the Fifth Fault (locality 80), producing a mean age of 1.75 ± 0.01 Myr (Table 1).

The time interval between Tuffs IB (1.80 Myr) and IF (1.75 Myr) is surprisingly brief, with most of the elapsed time

occurring in a narrow interval between Tuffs IB and IC (1.76 Myr). The most densely vegetated, wettest conditions, and highest lake levels of Bed I began just before Tuff IB and lasted to around the time of Tuff IC^{5,13}. A considerable climatic change occurred between Tuffs IC and IF, within the timespan of $\sim 10,000$ yr. The habitat became more open, and the climate became hotter and drier from Tuff IC to near the top of Bed I at FLK-N^{5,13,14}. Over the same time period, lake levels dropped

TABLE 2 Representative SCLF analyses

Naabi, OG89-56, Locality 66 ($J = 0.0002443 \pm 0.000001$)				
Ca/K	% Rad.*	Age (Myr)	$\pm$	
2486-02	0.0000	100.0	2.045	0.024
2486-03	0.0006	99.2	2.022	0.017
2486-04	0.0059	91.8	1.994	0.076
2486-05	0.0000	98.5	2.007	0.047
2486-06	0.0000	97.2	2.028	0.037
2486-07	0.0022	98.7	2.038	0.027
2486-08	0.0033	94.0	2.034	0.040
2486-09	0.0000	97.5	2.017	0.055
2486-10	0.0003	94.3	2.033	0.034
2486-01	0.0266	68.8	2.781	0.900
	0.0014		2.024 (mean)	0.005 (s.e.m.)
			2.029 (weighted mean)	0.005 (weighted uncertainty)
CFCT, OG-5, Locality 66b ($J = 0.001315 \pm 0.000005$)				
1505-03	0.0368	71.6	1.953	0.017
1505-04	0.0022	97.6	2.036	0.012
1505-06	0.0027	95.9	2.053	0.015
1505-07	0.0022	96.4	2.034	0.012
1505-08	0.1759	68.0	1.996	0.038
1505-09	0.0060	93.1	2.065	0.019
1505-10	0.0037	89.7	2.049	0.041
1505-01	0.1971	55.1	2.978	0.032
1505-02	0.9183	42.9	850.445	278.556
1505-05	8.2146	98.3	462.442	4.626
	0.0328		2.027 (mean)	0.015 (s.e.m.)
			2.030 (weighted mean)	0.014 (weighted uncertainty)
Tuff IA, OG-6, Locality 66a ($J = 0.0001622 \pm 0.0000002$)				
1841-06	0.4566	95.8	1.973	0.112
1841-04	0.5390	80.7	2.003	0.373
1841-03	0.0318	92.8	2.067	0.028
1841-02	8.8165	99.1	350.799	12.448
1841-09	6.4343	100.6	336.888	25.592
1841-05	0.0079	100.0	456.064	1.637
1841-01	0.0019	99.9	477.134	1.569
1841-07	9.0658	99.4	457.719	13.132
1841-08	0.0078	100.0	456.405	1.807
1841-10	23.2670	99.7	740.165	15.806
	0.4978		1.988 (mean)	0.015 (s.e.m.)
			1.976 (weighted mean)	0.108 (weighted uncertainty)
Tuff IB, OG-1, Locality 9 ($J = 0.001315 \pm 0.000005$)				
1507-06	0.0807	95.4	1.780	0.014
1507-07	0.0915	92.0	1.766	0.016
1507-08	0.0850	93.6	1.814	0.019
1507-09	0.0460	96.0	1.795	0.014
1507-10	0.0647	77.5	1.771	0.020
1507-11	0.0682	95.4	1.766	0.015
1507-12	0.0524	97.0	1.826	0.015
1507-13	0.0573	96.5	1.824	0.013
1507-14	0.0420	96.3	1.783	0.016
1507-15	0.1107	90.6	1.781	0.015
1507-16	0.0582	91.6	1.807	0.014
1507-17	0.0651	87.3	1.829	0.016
1507-18	0.0426	94.5	1.791	0.014
1507-19	0.0848	94.9	1.799	0.009
	0.0678		1.795 (mean)	0.006 (s.e.m.)
			1.797 (weighted mean)	0.006 (weighted uncertainty)

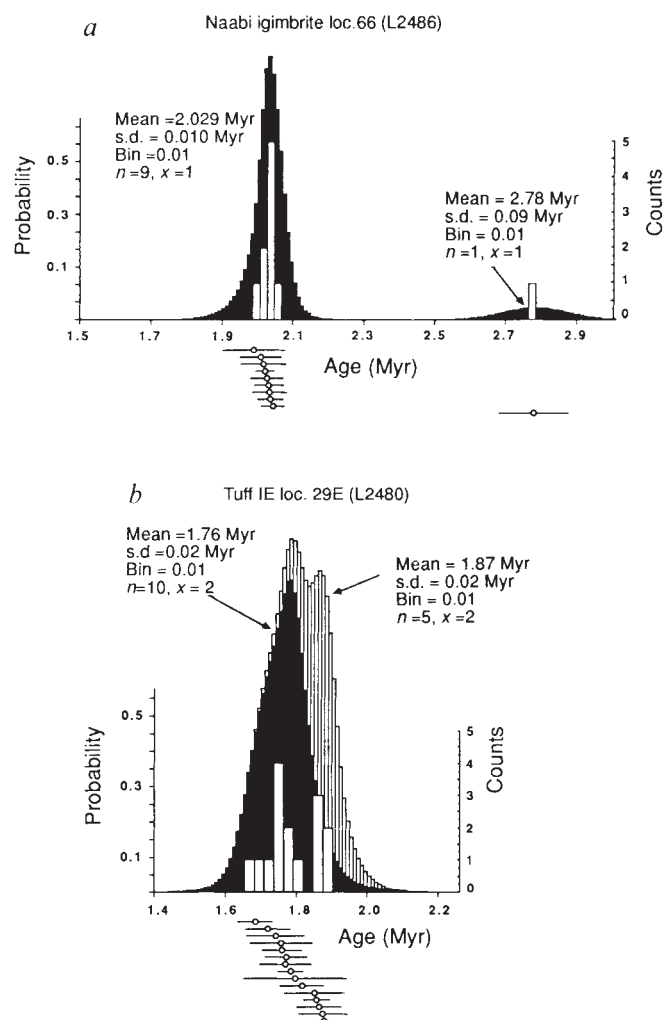


FIG. 2 Distribution plots of SCLF data for feldspars from the Naabi ignimbrite (a) and Tuff IE (b). Open bar chart symbols represent simple age-frequency histograms calculated with bin widths of 0.025, and with frequency (counts) along the right hand column. The dark or patterned curves superimposed on the histograms depict plots of probability-density functions for the same data set; here, a computer algorithm compiles the mean and standard deviation of each SCLF analysis, shown below the plot, to build a probability distribution based on gaussian statistics. The calculated age, shown beside each curve, is the weighted average and uncertainty²³ of the SCLF analyses. Bin, bin width; n , number of grains analysed; x is a scaling factor. Corroborative evidence for the age of the Naabi sandines is given by an Ar isotope correlation regression (Table 1). These results show that the initial $^{40}\text{Ar}/^{36}\text{Ar}$ value for the nine single population grains is 297 ± 64 , which indicates that the trapped argon component is atmospheric in composition, and that there is no systematic contribution from excess ^{40}Ar . The age calculated from the regression, 2.03 ± 0.01 Myr, is identical to the mean SCLF result. Furthermore, the MSWD value, which is less than 1, indicates that the scatter of data about the best-fit linear regression is purely random. These combined illustrations clearly resolve the contaminating grain from the primary age population, showing that the primary feldspar data are modelled well by simple random errors.

The neutron fluence monitor used for calculating J is the Fish Canyon sanidine, which has a reference age of 27.84 Myr²⁴. s.e.m., standard error of the mean²³. Errors in age ($\pm$) are 1σ analytical uncertainties. Weighted mean/weighted uncertainty are as in Table 1. Data in italics are excluded from calculation of mean age. Analytical methods used in this study are similar to those cited in Deino *et al.*²⁴.
* Radiogenic argon component. ($\lambda_e + \lambda_c = 0.581 \times 10^{-10} \text{ yr}^{-1}$; $\lambda_\beta = 4.962 \times 10^{-10} \text{ yr}^{-1}$; $^{40}\text{K}/\text{K} = 1.167 \times 10^{-4} \text{ mol mol}^{-1}$).

and the lake reduced in area. Shortly before the deposition of Tuff IF, lake levels rose again⁵, presumably owing to a more humid climate. The humid climate persisted into the lower part of Bed II below the Lemuta Member¹⁴. This rapid rate of climatic change is comparable to that observed in East African lakes during latest Pleistocene and Holocene times¹⁵.

Tuff IA falls within the Olduvai subchron and the base of the event probably occurs shortly below^{5,16} (Fig. 1), yet the currently accepted age for this boundary, 1.87 Myr¹², is based on K-Ar dates from volcanic units well above Tuff IA^{3,9,10,11}. The 1.98 Myr date for Tuff IA implies that the age of the lower boundary of the Olduvai subchron is between 1.98 and 2.01 Myr, and that a short normal event above the Naabi ignimbrite (2.03 Myr) and below CFCT (2.02 Myr) (Fig. 1, loc. 66b) is

probably the upper magnetozone of the Réunion subchron. If the age of Tuff IA is correct, then other palaeomagnetic chronologies based on the Olduvai subchron may need revision. This would alter the chronological interpretation of marine sequences, especially with respect to the global climatic record, near the Plio-Pleistocene boundary¹⁷⁻¹⁹.

It is stratigraphically permissible that 1.98 Myr is the eruptive age and the depositional age of Tuff IA, but the date itself is not well constrained (Table 2). There is evidence, however, from other sites in East Africa²⁰⁻²² supporting our observation that the base of the Olduvai subchron is older than 1.9 Myr. Nevertheless, given the broad implications of this result, we feel that further SCLF dating and a detailed magnetostratigraphic study of lower Bed I are warranted. □

Received 25 April; accepted 12 September 1991.

1. Leakey, L. S. B., Evernden, J. A. & Curtis, G. H. *Nature* **194**, 1-7 (1962).
2. Leakey, L. S. B., Evernden, J. A. & Curtis, G. H. *Nature* **191**, 478-479 (1961).
3. Evernden, J. A. & Curtis, G. H. *Curr. Anthropol.* **6**, 343-364 (1965).
4. Curtis, G. H. & Hay, R. L. in *Calibration of Hominoid Evolution* (eds Bishop, W. W. & Miller, J. A.) 289-302 (Scottish Academic Press, Edinburgh, 1972).
5. Hay, R. L. *Geology of the Olduvai Gorge: A Study of Sedimentation in a Semiarid Basin* 1-203 (University of California Press, Berkeley, 1976).
6. Steiger, R. H. & Jager, E. *Earth planet. Sci. Lett.* **36**, 359-362 (1977).
7. Lo Bello, P. et al. *Chem. Geol. (Isot. Geosci.)* **66**, 61-71 (1987).
8. Bogaard, P., v.d. Hall, C. M., Schminke, H.-U. & York, D. *Geophys. Res. Lett.* **14**, 1211-1214 (1987).
9. Grommé, C. S. & Hay, R. L. *Nature* **200**, 560-561 (1963).
10. Grommé, C. S. & Hay, R. L. *Earth planet. Sci. Lett.* **2**, 111-115 (1967).
11. Grommé, C. S. & Hay, R. L. *Earth planet. Sci. Lett.* **10**, 179-185 (1971).
12. Markinen, E. A. & Dalrymple, G. B. *J. geophys. Res.* **84**, 615-626 (1979).
13. Kappleman, J. *Quat. Res.* **25**, 141-149 (1986).
14. Bonnefille, R. *Nat. Geogr. Soc. Res. Rep.* **17**, 227-243 (1984).
15. Butzer, K. W., Isaac, G. L., Richardson, J. L. & Washburn-Kamau, C. *Science* **175**, 1069-1076 (1972).
16. Brock, A. & Hay, R. L. *Earth planet. Sci. Lett.* **29**, 126-130 (1976).
17. Shackleton, N. J. & Opdyke, N. D. *Geol. Soc. Am. Mem.* **145**, 449-464 (1976).

18. Tauxe, L., Opdyke, N. D., Pasini, G. & Elmi, C. *Nature* **304**, 125-129 (1983).
19. Berggren, W. A., Kent, D. V. & van Couvering, J. A. in *The Chronology of the Geological Record* (ed. Snelling, N. J.) 211-260 (Blackwell Scientific Publications, Oxford, 1985).
20. McDougall, I. *Geol. Soc. Am. Bull.* **96**, 159-175 (1985).
21. Brown, F. H., Shuey, R. T. & Croes, M. K. *Geophys. J. R. astr. Soc.* **54**, 519-538 (1978).
22. Hillhouse, J. W., Cerling, T. E. & Brown, F. H. *J. geophys. Res.* **91**, 11581-11595 (1986).
23. Taylor, J. R. *An Introduction to Error Analysis: The Study of Uncertainty in Physical Measurements* 1-270 (University Science, Mill Valley, CA 1982).
24. Deino, A., Tauxe, L., Monaghan, M. & Drake, R. *J. Geol.* **98**, 567-587 (1990).
25. Samson, S. D. & Alexander, E. C. *Chem. Geol. (Isot. Geosci.)* **66**, 27-34 (1987).
26. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* 1-336 (McGraw-Hill, New York, 1969).
27. Kauffman, D. S., Walter, R. C., Brigham-Grette, J. & Hopkins, D. M. *Quat. Res.* (in the press).
28. Leakey, M. D. *Excavations in Beds I & II 1960-1963* 1-306 (Cambridge University Press, 1971).

ACKNOWLEDGEMENTS. We acknowledge S. Waane, Director of the Tanzanian Department of Antiquities, and Professor Msangi, Director General of the Tanzanian Commission for Science and Technology, for allowing this project to occur. We thank G. Eck, D. Johanson, W. Kimbel, F. Masao, P. Ndessokia, and T. White for their contributions to the project, and F. C. Howell, W. Kimbel and G. Eck for comments on the manuscript. We thank S. Shute for mineral separations, and D. Beckner for graphics. The research was partially funded by the L. S. B. Leakey Foundation, the NSF and David Koch.

A primary determinant for lipoygenase positional specificity

David L. Sloane*, Regina Leung*, Charles S. Craik† & Elliott Sigal*‡§

* Cardiovascular Research Institute, and Departments of † Pharmaceutical Chemistry and ‡ Medicine, University of California, San Francisco, California 94143-0911, USA

THE three mammalian lipoygenases are named according to the carbon position (5, 12 or 15) at which they catalyse the oxygenation of arachidonic acid¹; they are implicated in inflammatory disorders, for example 15-lipoygenase is induced in atherosclerosis² and can oxidize low-density lipoprotein to its atherogenic form^{3,4}. To identify what determines this positional specificity, we have exchanged conserved differences in the isoforms of 12- and 15-lipoygenases. Substitution of methionine with valine at position 418 of human 15-lipoygenase results in an enzyme that performs 12- and 15-lipoygenation equally. This effect can be mimicked by incubating wild-type 15-lipoygenase with a synthetically altered substrate which has its doubly allylic methylene carbons shifted by one carbon relative to arachidonic acid. Other mutations at the neighbouring amino acids 416 and 417 give an enzyme which performs 12- and 15-lipoygenation in a ratio of 15:1. These results indicate that this region might position the substrate in the active site.

At present we have no high-resolution three-dimensional structure for any lipoygenase, and as the lipoygenases do not share extensive homology with any other proteins, computer-aided modelling of critical areas of the enzyme is not feasible.

We have therefore relied on the identity and variation in sequence alignments of the lipoygenases to locate amino-acid residues essential for catalysis and substrate specificity. The lipoygenases constitute a family of highly homologous enzymes: isoforms of 12- and 15-lipoygenase are the most similar (65-90%) but there is no apparent pattern of significant differences. Among the sequences of the three 12-lipoygenases however, 414 out of 661 amino acids are identical⁵⁻⁷. Of these, 47 amino acids differ from their counterparts in 15-lipoygenase. In this set of 47 differences between the 12- and 15-lipoygenases, only four amino acids are conserved between the two known 15-lipoygenase sequences^{8,9} (Fig. 1). These four conserved differences, or a subset of them, are candidates for positioning the substrate relative to the active site and hence may determine either 12- or 15-lipoygenation.

When these four amino acids in human 15-lipoygenase are changed to amino acids found in 12-lipoygenase, the enzyme performs 12- and 15-lipoygenation equally (Table 1). Analysis of enzymes with single, double, triple and quadruple mutations showed that Met 418 is the primary determinant of positional specificity. When Met 418 is replaced with a Val (15-LOX, M418V in one-letter amino-acid code), either by itself or in combination with other changes, the resulting enzyme catalyses 12- and 15-lipoygenation equally, but not 5-lipoygenation (Fig. 2a, b). None of the other amino-acid changes is responsible for the positional specificity change (Table 1).

The resultant change in specificity of the M418V variant implies a shift in the binding of the substrate in the active site. This shift may indicate that the atoms of the side chain at position 418 are in contact with the substrate. Replacement of a methionine with a valine results in a 17% reduction of side-chain volume¹⁰. This would lead to a shift in substrate binding if the atoms of this side chain were to interact with the substrate. As Met and Val are both hydrophobic amino acids, they could participate in binding a fatty acid substrate.

§ To whom correspondence should be addressed.

FIG. 1 Sequence comparisons of the mammalian lipoxygenases. The regions containing the four conserved changes are shown. The sequences are the human reticulocyte 15-lipoxygenase (hum15LO)⁸, rabbit reticulocyte 15-lipoxygenase (rab15LO)⁹, bovine tracheal epithelial 12-lipoxygenase (bov12LO)⁷, porcine leukocyte 12-lipoxygenase (porc12LO)⁵, human platelet 12-lipoxygenase (hum12LO)⁶, human leukocyte 5-lipoxygenase (hum5LO)^{25,26} and rat leukocyte 5-lipoxygenase (rat5LO)²⁷. Numbers flanking each sequence correspond to the number of the first and last amino acid in that region, where the initial methionine of the protein is number 1. A dash indicates amino-acid identity with human reticulocyte 15-lipoxygenase. The single-letter amino-acid code is used. Arrows indicate the site of the conserved differences between 15- and 12-lipoxygenases. For completeness, sequences of the 5-lipoxygenases in these regions are shown. There are additional conserved differences between 5-lipoxygenase and the other lipoxygenase sequences.

hum15LO : 100	EGNGVLSLPE	109	hum15LO : 392	LRYTLEINVR	401
rab15LO : 102	V-D--Q---V	111	rab15LO : 394	-----	403
bov12LO : 102	--D-I-----	111	bov12LO : 394	---M---I-	403
porc12LO : 102	--DRI-----	111	porc12LO : 394	F---M-----	403
hum12LO : 101	Q-EDI-----	110	hum12LO : 393	I---M---T-	402
hum5LO : 105	T-DVEVVLRD	114	hum5LO : 401	V-F-IA--TK	410
rat5LO : 104	T-E-EIVLDR	113	rat5LO : 400	V-F-IA--TK	409

hum15LO : 414	FDQIMSTGGGGHV	426	hum15LO : 427	QLLKQAGAF	436
rab15LO : 415	-----	427	rab15LO : 428	---Q-----R	437
bov12LO : 415	---VV-----	427	bov12LO : 428	E--QR-----	437
porc12LO : 415	---VV-----	427	porc12LO : 428	E--RR-A-L-	437
hum12LO : 414	--KAV-----	426	hum12LO : 427	---RR-A-Q-	436
hum5LO : 422	--KANA-----	434	hum5LO : 435	-MVQR-MKD-	444
rat5LO : 421	--KANA-----	433	rat5LO : 434	-MVQR-VQD-	443

The 15-lipoxygenase from rabbit reticulocytes shows a comparable degree of dual positional specificity¹¹ on the substrate 6,9,12,15-20:4 (ω 5) eicosatetraenoic acid. The doubly allylic methylene carbons on this substrate are shifted by one carbon relative to arachidonic acid (5,8,11,14-20:4(ω 6) eicosatetraenoic acid). The ω 5 substrate presumably repositions the reactive doubly allylic methylenes in the active site of the enzyme¹¹. Likewise, the altered regio-specificity of 15-LOX M418V suggests that arachidonic acid binds to the enzyme so that the carboxyl end of the substrate is closer to the active site. To test this, we analysed the activity of wild-type and 15-LOX M418V enzymes with arachidonic acid and with the ω 5 eicosatetraenoic acid. The reaction mechanism of 15-lipoxygenase is not precisely known, but the initial step is probably a hydrogen abstraction at the methylene carbon 13 (C13), followed by a double-bond rearrangement, leading to oxygen insertion at C15 (ref. 12). On arachidonic acid, the site responsible for hydrogen abstraction of the wild-type enzyme is close to carbon 13, leading predominantly to 15-lipoxygenation (Fig. 2a). But on the ω 5 substrate, C13 is no longer a methylene and

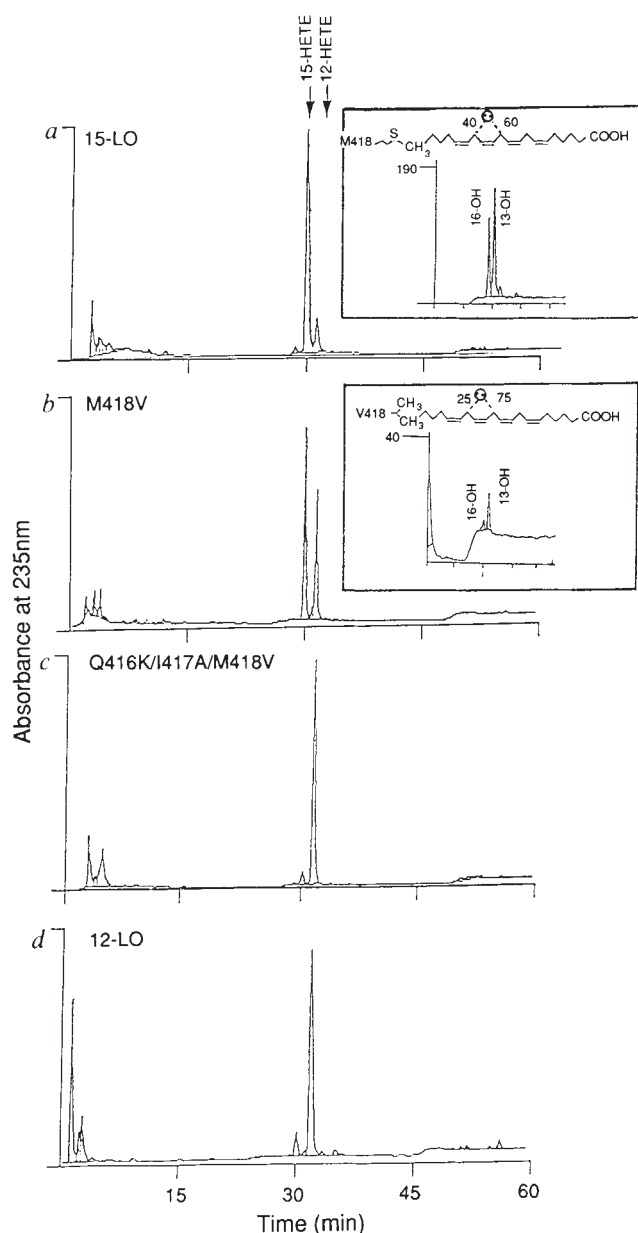


FIG. 2 Reverse-phase HPLC profiles (monitored at 235 nm) of the oxygenation products of wild-type and mutant lipoxygenases. Assay conditions are as described in Table 1. Wild-type human reticulocyte 15-lipoxygenase (LO) was incubated with 300 μ M 5,8,11,14-20:4 (ω 6) eicosatetraenoic acid (arachidonic acid) (a) or with 60 μ M 6,9,12,15-20:4 (ω 5) eicosatetraenoic acid (a, inset). Mutant human reticulocyte lipoxygenase (M418V) was incubated with 300 μ M 5,8,11,14-20:4 (ω 6) (b) or with 60 μ M 6,8,12,15-20:4 (ω 5) (b, inset). Above the insets is shown the proposed interaction of residue 418 with the methyl end of the substrate. Shaded circles represent the hydrogen acceptor centre in the enzymes. Numbers on dashed lines indicate the relative amounts of hydrogen abstraction from each carbon, as calculated from the relative amounts of product. The activities of the wild-type and mutant enzymes on arachidonic acid are equal. On the 6,9,12,15-20:4 (ω 5) eicosatetraenoic acid substrate, the wild-type enzyme is 30% as active as it is on arachidonic acid. The M418V mutant enzyme is less than 0.1% as active on 6,9,12,15-20:4 (ω 5) eicosatetraenoic acid as on arachidonic acid, yielding no product in four out of six reactions. The product profile of the triple mutant, 15LOX Q416K/I417A/M418V, incubated with 300 μ M arachidonic acid, is shown in c. This variant was made using a 30-mer that contained seven mismatches to convert the codons CAG (Gln 416) to AAA (Lys), ATA (Ile 417) to CGT (Ala), and ATG (Met 418) to GTT (Val), as described in Table 1. The products formed by this variant lipoxygenase are in a ratio of 15:1 (12-HETE:15-HETE) and resembles that of wild-type bovine tracheal epithelial 12-lipoxygenase (d). Bovine 12-lipoxygenase was assayed by incubating 300 μ M arachidonic acid with *E. coli* transformed with a plasmid containing the 12-lipoxygenase complementary DNA from bovine tracheal epithelial cells⁷. Peaks are labelled according to co-elution with standards.

the hydrogen abstraction must occur at C11 or C14 (which are equidistant from the reaction centre), leading to equal 13- and 16-hydroxy products (Fig. 2a, inset). Presumably, the 15-LOX M418V enzyme has a deeper substrate-binding pocket than the wild-type enzyme, so the C13 and C10 of bound arachidonic acid are equidistant from the reaction centre, leading to equal 15- and 12-lipoxygenation (Fig. 2b). Therefore the mutant enzyme should predominantly oxygenate C13 of the ω 5 substrate and, as expected, the reaction yields mainly 13-hydroxy-eicosatetraenoic acid (Fig. 2b, inset), although the ω 5 substrate is a very poor substrate for the mutant enzyme (<0.1% of wild-type).

The region immediately surrounding amino-acid 418 is highly conserved in all mammalian lipoxygenases (nine of thirteen residues are identical; Fig. 1). Because the human platelet 12-lipoxygenase is more stringent in its product formation performing exclusively 12-lipoxygenation⁶, we mutated 15LOX M418V further to contain the platelet 12-lipoxygenase amino acids in this region. This variant, 15-LOX Q416K/I417A/M418V, generated 12-lipoxygenase and 15-lipoxygenase products in a ratio of 15:1 (Fig. 2c), like bovine epithelial 12-lipoxygenase (Fig. 2d).

The wild-type and mutant enzymes were tested for their stereospecificity and specific activity. The products of the wild-type and mutant enzymes were analysed for chirality by HPLC and found to be the S-isomers (S:R ratio >95%), indicating that they were formed enzymatically. The levels of expression of wild-type and variant lipoxygenases are <0.1% of the total

TABLE 1 Product formed by *E. coli* cells carrying the wild-type and mutant lipoxygenases

Mutant	Product formation (pmol per 15 min)		
	15-HETE	12-HETE	Ratio
Wild-type (15-LO)	6,780	630	1:9
V104I, L397M, M418V, Q431R	3,660	3,220	1:1
L397M, M418V, Q431R	2,690	3,000	1:1
L397M, M418V	3,340	3,090	1:1
V104I	2,630	280	1:9
L397M	3,310	340	1:9
M418V	4,030	4,250	1:1
Q431R	4,500	485	1:9

The various mutant enzymes were analysed for the production of hydroxyeicosatetraenoic acids (HETEs) after incubation with exogenous arachidonic acid. Site-directed mutagenesis was performed by the method of Kunkel²² using mismatched synthetic oligonucleotides to mutate the codons. For the V104I mutation (Val at position 104 in the primary sequence of 15-lipoxygenase (15-LO) mutated to Ile), a 20-mer was used that contained two mismatches to convert the GTC (Val 104) codon to ATA (Ile). Uracil-laden single-stranded DNA was prepared from this mutant to generate the quadruple mutant, using a 120-mer that contained three mismatches to convert the CTG (Leu 397) to an ATG (Met), the ATG (Met 418) to a GTG (Val) and the CAA (Gln 431) to a CGA (Arg). The triple mutant was made using this 120-mer on wild-type, uracil-laden single-stranded DNA. The L397M/M418V double mutant and the M418V single mutant were recovered using the 120-mer as a mismatched primer. The L397M and Q431R single mutants were made using 22-mer oligonucleotides which each contained single mismatches. Uracil-laden plasmid pSS15LO was used as a template. This plasmid contains a *lac* promoter to express the 15-lipoxygenase and a bacteriophage f1 origin of replication to generate single strands. After the mutagenesis reaction, transformants were screened by digestion with restriction enzymes. Positive mutants were verified by DNA sequencing²³. The complete coding region of the gene containing the M418V mutation was sequenced to ensure that there were no second-site mutations. Enzyme activity was assayed as described⁷. Briefly, arachidonic acid (300 μ M) was incubated (at 37 °C for 15 min) with the *E. coli* whole cells containing the mutation. Cells were sonicated, debris removed by centrifugation and the lipids extracted and analysed by reverse-phase high-pressure liquid chromatography²⁴. HETEs were quantitated using an extinction coefficient of 23,500 (mol per litre)⁻¹ cm⁻¹ and prostaglandin B₂ as an internal standard. Values are from representative reactions and are reproducible.

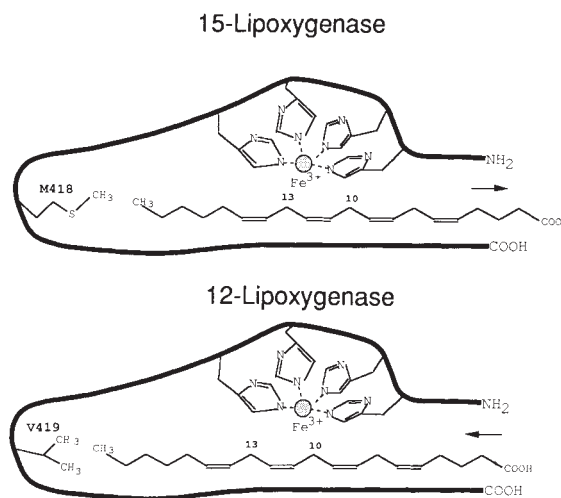


FIG. 3 Model depicting the difference in substrate positioning by 12- and 15-lipoxygenases. The substrate-binding cleft and a potential iron-binding site are shown. At the bottom of the binding cleft is amino acid 418 (419 in 12-lipoxygenase). The smaller size of this side chain in 12-lipoxygenases allows deeper binding of the substrate. This shifts the substrate relative to the site of hydrogen abstraction (here depicted as the iron). The iron is shown chelated by histidines, which is consistent with spectroscopic evidence from soybean lipoxygenases^{28,29} and mutagenesis data on human lipoxygenases.

Escherichia coli cellular proteins¹³, so we have estimated the relative specific activities of the wild-type and variant lipoxygenases by electrophoresis of a serial dilution of bacterial cultures and detection of lipoxygenase by immunoblotting. Expression of wild-type and mutant lipoxygenases is comparable (data not shown), as is the total activity (15- plus 12-lipoxygenation) (Table 1). We conclude that the specific activity is the same for the mutants and the wild type.

Although the active-site residues of lipoxygenase have not yet been identified, amino acid 418 is adjacent to a region containing five histidines which are conserved in all known lipoxygenases (in 15-lipoxygenase, amino acids 355, 360, 365, 383 and 392) and are believed to be important in binding the catalytically active iron moiety^{14,15}. Mutagenesis of these histidines has shown that a subset is essential for catalysis (D.L.S., unpublished results; and ref. 16). Because Met 418 plays a part in positioning the substrate and is close to the histidine-rich region, our results help to define further the active-site binding cleft of the enzyme and suggest a model for enzyme-substrate interaction (Fig. 3). It has been previously shown that the specificity of 15-lipoxygenase is affected by the distance of a doubly allylic methylene carbon from the methyl end of the fatty acid substrate^{12,17}, but the amino acids responsible for this positional specificity were unknown. In our model, amino acid 418 (in 15-lipoxygenase) or 419 (in 12-lipoxygenase) interacts with the methyl terminal moiety of the substrate. The side-chain bulk of this residue shifts the substrate relative to the reaction centre so that different methylene carbons are aligned with the active site. Deeper binding of the methyl end of the substrate shifts the site of hydrogen abstraction from C13 to C10 of arachidonic acid, leading to 12- rather than 15-lipoxygenation. Our results also indicate that amino acids adjacent to Met 418 contribute to this enzyme-substrate interaction.

There are examples of engineered changes in substrate specificity of enzymes¹⁸⁻²¹, but all have relied on the three-dimensional structure of the enzyme. We have presented an example of engineering a positional specificity change without a three-dimensional structure, an approach that could be applied to any closely related family of enzymes to give structure-function information, especially now that sequences of

presumed active site regions can be generated by the polymerase chain reaction. □

Received 26 July; accepted 6 September 1991.

- Samuelson, B. *et al. Science* **237**, 1171–1176 (1987).
- Yla-Herttuala, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 6959–6963 (1990).
- Sparrow, C. P. *et al. J. Lipid Res.* **29**, 745–753 (1988).
- Steinberg, D. *et al. New Engl. J. Med.* **320**, 915–924 (1989).
- Yoshimoto, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 2142–2146 (1990).
- Funk, C. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 5638–5642 (1990).
- De Marzo, N. *et al. Am. J. Physiol.* (in the press).
- Sigal, E. *et al. Biochem. biophys. Res. Commun.* **157**, 457–464 (1988).
- Fleming, J. *et al. Gene* **79**, 181–188 (1989).
- Ponder, J. W. & Richards, F. M. *J. molec. Biol.* **193**, 775–791 (1987).
- Kuhn, H. *et al. J. biol. Chem.* **265**, 16300–16305 (1990).
- Hamberg, M. & Samuelsson, B. *J. biol. Chem.* **242**, 5329–5335 (1967).
- Sloane, D. L. *et al. Biomed. biochim. Acta* **49**, S11–S16 (1990).
- Pistorius, E. K. & Axelrod, B. *J. biol. Chem.* **249**, 3183–3186 (1974).
- Shibata, D. *et al. J. biol. Chem.* **263**, 6816–6821 (1988).
- Nguyen, T. *et al. J. biol. Chem.* (in the press).
- Bryant, R. W. *et al. J. biol. Chem.* **257**, 6050–6055 (1982).
- Craig, C. S. *et al. Science* **228**, 291–297 (1985).
- Fersht, A. R. *et al. Nature* **314**, 235–238 (1985).
- Wells, J. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5167–5171 (1987).
- Wilks, H. M. *et al. Science* **242**, 1541–1544 (1988).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488–492 (1985).
- Sanger, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Sigal, E. *et al. J. biol. Chem.* **265**, 5113–5120 (1990).
- Matsumoto, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 26–30 (1988).
- Dixon, R. A. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 416–420 (1988).
- Balcerek, J. M. *et al. J. biol. Chem.* **263**, 13937–13941 (1988).
- Navaratnam, S. *et al. Biochim. biophys. Acta* **956**, 70–76 (1988).
- Dunham, W. R. *et al. Eur. J. Biochem.* **190**, 611–617 (1990).

ACKNOWLEDGEMENTS. We thank H. Kuhn for discussion and A. Brash and H. Sprecher for the $\omega 5$ substrate. This work was supported in part by the NIH, the NSF (C.S.C.), the Council for Tobacco Research, and the Parke-Davis Pharmaceutical Research Division, Warner-Lambert Co. D.L.S. is a recipient of an American Heart Association research fellowship and E.S. is an Arthritis Foundation Investigator.

1b). After the delay period (4 s), a choice of two stimuli, the paired associate of the cue and one from a different pair, was shown. The monkey obtained fruit juice as a reward for correctly touching the paired associate within 1.2 s. In the recording sessions after training (Fig. 1 legend), the monkeys' performance was 70–100% correct. Extracellular spike discharges of single neurons were recorded from the anterior part of the temporal cortex (Fig. 2a), as reported in previous studies^{8,9}.

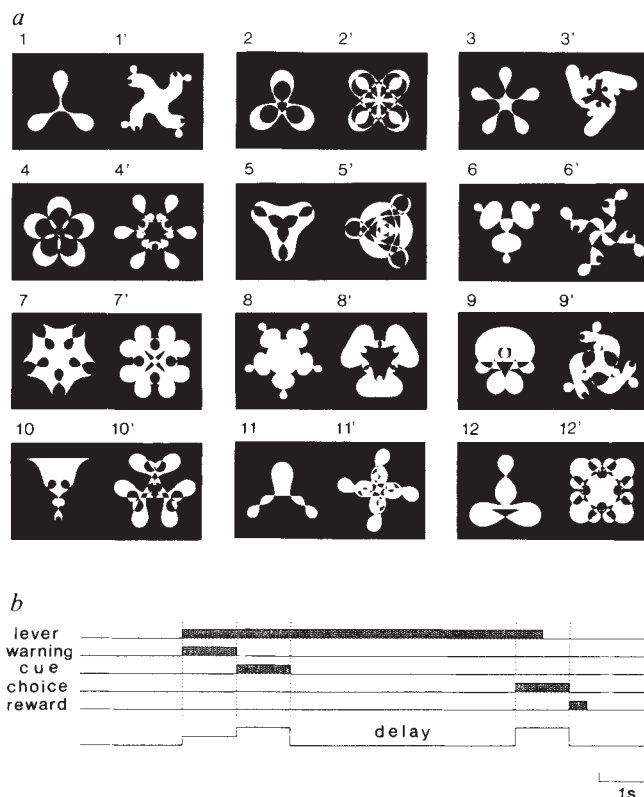


FIG. 1 Pair-association task for monkeys used to assess long-term memory. *a*, 12 pairs of Fourier descriptors (1 and 1'–12 and 12') for stimuli in the task. The monkeys learned to retrieve the other member of the pair associated with each cue picture. *b*, Sequence of events in a trial. Lever, lever press by the monkey to initiate a new trial; warning, green square (1 s); cue, one of 24 pictures as a cue stimulus (1 s); delay, green square (4 s); choice, a choice of two stimuli, the paired associate of the cue and one from a different pair; reward, fruit-juice reward for correctly touching the associate. Bottom trace, events chart used in Figs 2b, c and 3.

METHODS. Fourier descriptors were generated according to the reconstruction theorem with specified sets of harmonic amplitudes and phase angles²³. The real images on a video monitor were yellow monochrome against a black field. This set was used for one monkey, and coloured fractal patterns⁸ were used for the other. Sorting the pictures into pairs is basically random, avoiding apparent geometrical resemblances such as rotational symmetry. In each trial, cue stimulus was presented in the centre of the monitor screen and stimuli for choice were shown randomly in two of four positions (arranged in two rows of two columns). If the monkey released the lever before the choice, that trial was aborted. Error trials were not included in the rastergrams of Figs 2b, c and 3. The collected trials shown in these figures were originally separated by intervening trials of other cue stimuli, and were sorted by off-line computation. Training procedures of this pair-association task are as follows. Twelve pairs used in neuronal recordings were divided into three blocks (four pairs per block). After separate learning of three blocks, all 12 pairs were presented in random order. The direction of association (for example from 1 to 1', or from 1' to 1) was randomized except for the early training phase. Delay interval was 0.5 s before thorough randomization, and was increased to 4 s. Error correction trials were included in the early training phase. The criterion for acquisition was two consecutive days of 26 correct responses in 30 trials (87%). The two monkeys took 876 ± 303 trials per picture (mean $\pm$ s.e.m.) to reach this criterion. Extracellular spike discharges of single neurons were recorded using standard physiological techniques²⁴.

Neural organization for the long-term memory of paired associates

Kuniyoshi Sakai & Yasushi Miyashita

Department of Physiology, School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

MOST of our long-term memories of episodes or objects are organized so that we can retrieve them by association. Clinical neuropsychologists assess human memory by the paired-associate learning test, in which a series of paired words or figures is presented and the subject is then asked to retrieve the other pair member associated with each cue¹. Patients with lesions of the temporal lobe show marked impairment in this test^{2–6}. In our study, we trained monkeys in a pair-association task⁷ using a set of computer-generated paired patterns. We found two types of task-related neurons in the anterior temporal cortex. One type selectively responded to both pictures of the paired associates. The other type, which had the strongest response to one picture during the cue presentation, exhibited increasing activity during the delay period when the associate of that picture was used as a cue. These results provide new evidence that single neurons acquire selectivity for visual patterns through associative learning. They also indicate neural mechanisms for storage and retrieval in the long-term memory of paired associates.

We prepared 24 computer-generated pictures for each monkey, and sorted geometrically distinct patterns into pairs (Fig. 1a). The combination of the paired associates is not predictable without memorizing them beforehand. Two macaque monkeys (*Macaca fuscata*) were trained to memorize a set of 12 pairs through repeated trials in the pair-association task. In each trial, a cue stimulus was presented on a video monitor for 1 s (Fig.

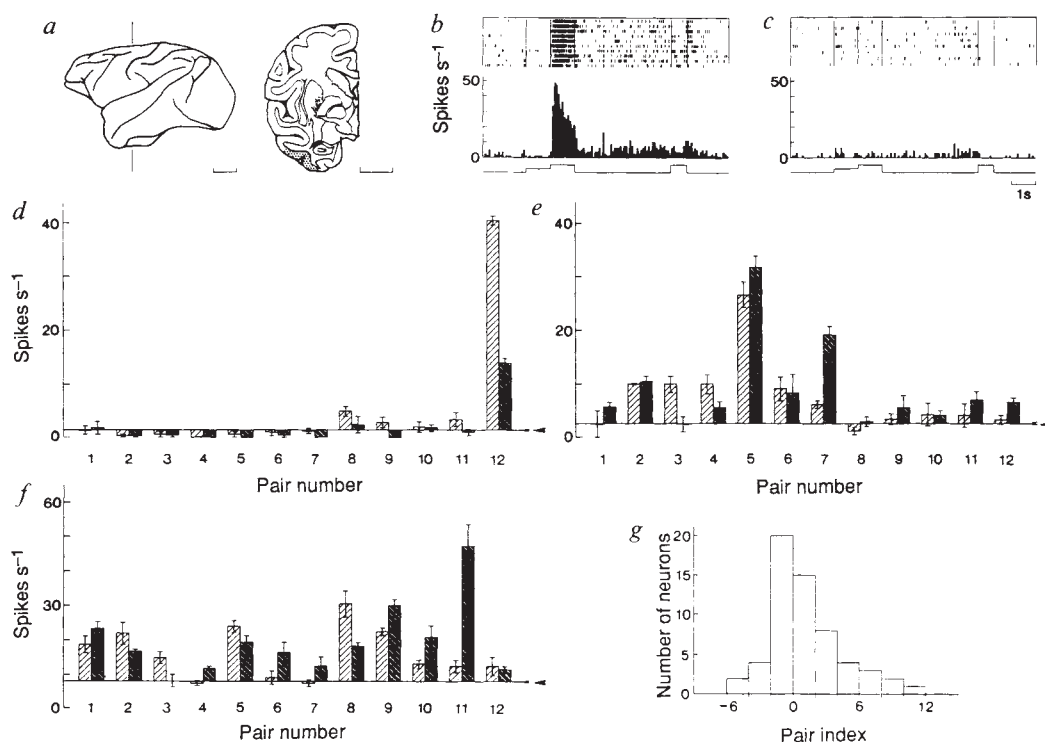
Figure 2 shows one type of neuron with picture-selective responses during the cue period. One picture elicited the strongest response during the cue period from a neuron, with some activity during the delay period (Fig. 2*b*). By contrast, another picture elicited no response at all (Fig. 2*c*). This neuron responded reproducibly to only a few pictures, but not to other pictures in the set. It might be that the cell responded to geometrically similar patterns. The strongest and the second-strongest responses were ascribed to a particular pair which had no apparent geometrical similarity (Fig. 2*d*). Some other cells showed broader tuning and responded to more than three pictures. Nevertheless, paired pictures were found to be among the most effective stimuli for these cells (Fig. 2*e, f*). We call this type of cell a 'pair-coding neuron', which manifests selective cue responses to both pictures of the paired associates.

Of 577 isolated neurons, 91 cells reproducibly showed a strong picture-selective response during the cue period (Fig. 2 legend). The most effective stimuli for the 91 cells covered all pictures in the set. These responsive cells tended to be located near to

one another (1–2 mm wide) in the temporal cortex. Thirty-two of the 91 cells responded to only one picture, whereas 59 cells responded to more than two pictures. We further analysed these 59 cells by calculating two coupling indices for each neuron (see legend to Fig. 2). One coupling index (denoted as CI_p) measures correlated neural responses to paired associates, whereas the other coupling index (CI_r) estimates responses to other random combinations among 24 pictures. The latter index CI_r serves as an experimental control for untrained association between two pictures. For each cell, we defined a pair index (PI) as equal to $CI_p - CI_r$. The frequency distribution of PI values is shown in Fig. 2*g*, demonstrating that the paired associates elicited significantly correlated responses ($P < 0.015$; Wilcoxon's signed-rank test, $n = 59$). We conclude that the selectivity of these neurons was acquired through learning of the pair-association task.

We found another type of neuron with picture-selective activities during the delay period. One picture elicited the strongest response during the cue period from a single neuron

FIG. 2 Responses of 'pair-coding neurons' which were selective to both pictures of the paired associates. *a*, Location of recorded neurons. Left, lateral view of a monkey brain. Right, cross-section indicated by a vertical line on the lateral view. The stippled area represents the range of recording sites. Scale bars, 10 mm. *b*, Rastergrams of neural discharges in each trial (upper) and spike-density histograms (lower) obtained from a single neuron. Bin width, 80 ms. These trials were collected for cue 12 which elicited the strongest response during the cue period. *c*, Trials for cue 7 which elicited no response at all in the same cell as *b*. *d*, Mean discharge rates for each cue presentation (mean $\pm$ s.e.m.) relative to the spontaneous discharge rate (denoted by an arrowhead) in the same cell as *b* and *c*. Cue stimuli are labelled as 'Pair number' on the abscissa (light histogram bar in number 1: cue 1, dark histogram bar in number 1: cue 1' and so on). This neuron selectively responded to both pictures of the paired associates 12 & 12'. *e*, Mean discharge rates for another cell. This neuron responded optimally to both pictures of the paired associates 5 & 5'. *f*, Mean discharge rates for another cell, whose pair index (see below) was equal to the mean value (1.3). In the cells shown in *d* and *e*, pair indices were 8.7 and 6.5, respectively. *g*, Frequency histogram showing the distribution of the pair indices in 59 responsive neurons in two monkeys. Positive values indicate that the neurons exhibited more correlated responses to paired associates than other random combinations, whereas negative values indicate the converse.



with $j \neq i'$ for random combinations, where x_i denotes a mean discharge rate during the cue period for the i th picture (the i th and i' th pictures belong to a pair), b is a spontaneous discharge rate, x_{best} and $x_{2\text{nd-best}}$ are mean discharge rates for the best and second-best cue-optimal pictures in each cell, N_p and N_r are the total number of combinations for two cases. Pair index (PI) was then defined as:

$$PI = CI_p - CI_r.$$

Evaluation of cue responses was done by calculating a mean discharge rate for each picture as follows. Spike numbers were collected over 400 ms at the beginning of the cue interval. They were averaged across trials for the same cue stimulus and their variances were evaluated to test reproducibility in each cell. Out of 577 isolated neurons, 436 cells were unresponsive. The pair index of weakly responsive cells is very susceptible to random fluctuations about a spontaneous discharge level. Moreover, weak responses could not be ascribed to the optimal stimulus²⁵. We therefore examined 104 cells whose discharge rates ($x_{\text{best}} - b$) were distributed beyond 15.5 Hz (the leftmost saddle point in the distribution of 141 responsive cells). Out of the 104 cells, 13 showed nonselective responses to all pictures, whereas the other 91 cells were picture-selective.

METHODS. We defined two coupling indices as:

$$CI_p = N_p^{-1} \sum_{i < j} \frac{(x_i - b)(x_j - b)}{(x_{\text{best}} - b)(x_{2\text{nd-best}} - b)} \times 100,$$

with $j = i'$ for paired associates,

$$CI_r = N_r^{-1} \sum_{i < j} \frac{(x_i - b)(x_j - b)}{(x_{\text{best}} - b)(x_{2\text{nd-best}} - b)} \times 100,$$

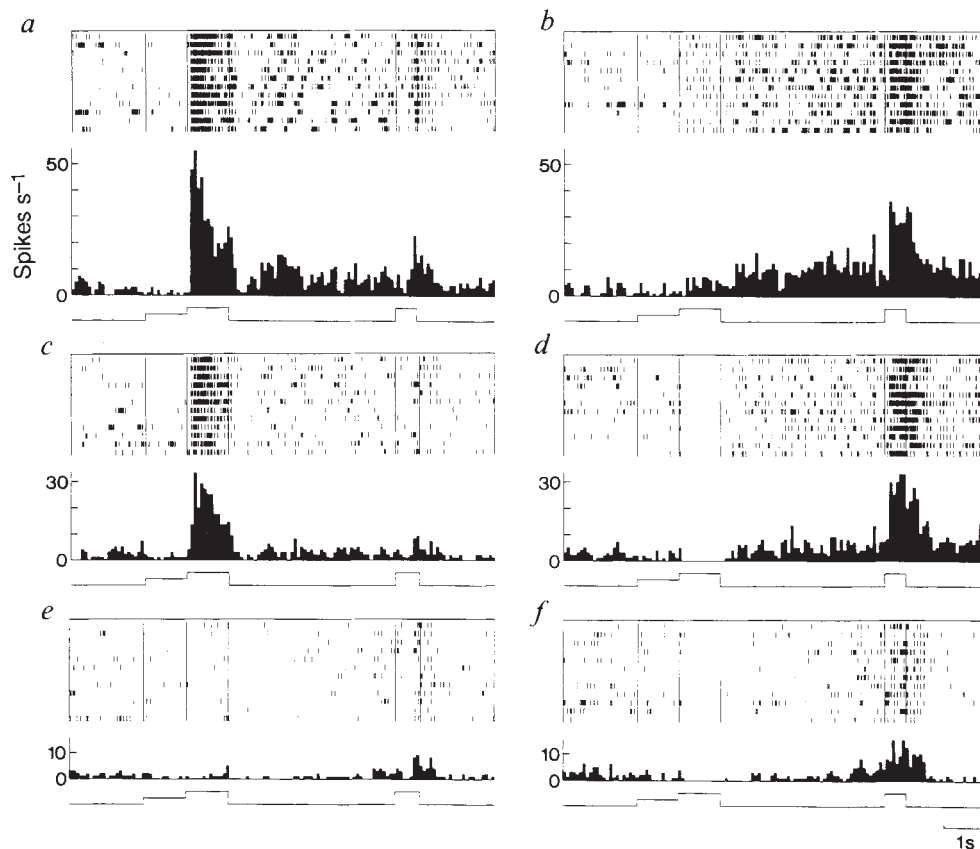


FIG. 3 Responses of a 'pair-recall neuron' which exhibited picture-selective activity during the delay period, presumably reflecting retrieval of the paired associate. *a-f*, Rastergrams and spike-density histograms obtained from a neuron. Bin width, 80 ms. *a*, Trials for cue 12 which elicited the strongest cue response. *b*, Trials for cue 12'. Note the tonic increasing activity during the delay period, which is much higher than the cue response. *c*, Trials for cue 1 that resulted in the second-strongest cue response. *d*, Trials for cue 1'. Note the sustained delay activity and the inhibitory cue response. *e*, Trials for cue 3 that elicited no response. *f*, Trials for cue 3'. In *f*, the discharges during the choice period are due to stimuli for choice other than 3, such as 1.

(Fig. 3*a*). In the trial when the paired associate of this cue-optimal picture was used as a cue, the same cell had the highest tonic activity during the delay period in contrast to a weak response during the cue period (Fig. 3*b*). This delay activity gradually increased until the choice of stimuli appeared. Furthermore, the paired associate of the second-best cue-optimal picture still elicited a sustained activity during the delay period (Fig. 3*c, d*). Other pictures evoked weak or no response (Fig. 3*e, f*). The delay activities were confined to a few cue stimuli in the set. We call this type of cell a 'pair-recall neuron', in which the paired associate of a cue-optimal picture elicited the highest delay activity.

Eleven of 91 cells showed picture-selective delay activities that surpassed cue responses in those trials. Out of 11 cells, 10 were pair-recall neurons as defined above. The highest delay activity of the pair-recall neurons does not represent mere sensory after-discharge, because it is stronger than the cue response. Furthermore, a significant augmentation of discharge rates was observed for the highest delay activity when mean discharge rates at two intervals were compared: 200–1,400 ms (near the beginning of the delay interval) and 2,760–3,960 ms (near the end) after delay onset ($P < 0.05$; Wilcoxon's signed-rank test, $n = 11$). By contrast, the delay activity elicited by a cue-optimal picture itself was significantly reduced during the delay period ($P < 0.005$; $n = 11$).

Out of 91 responsive neurons, 18 cells showed more than three error trials (where the monkey made the incorrect choice) in which a cue-optimal picture was presented. There was no significant difference in mean discharge rates during the cue period between correct and error trials ($P > 0.05$; Wilcoxon's signed-rank test, $n = 18$). Out of 10 pair-recall neurons, two cells exhibited more than three error trials in which the paired associate of a cue-optimal picture was presented. One cell showed no significant difference, whereas the other cell showed a higher delay activity in correct trials than in error trials, which correlated with the monkey's choice ($P < 0.01$; two-tailed modified t test¹⁰, $t' = 3.21$ with 17 d.f.).

We previously reported that a picture-selective delay activity reflected stimulus-stimulus association caused by the fixed order of picture presentation in a visual delayed matching-to-sample task⁹. But evidence from that experiment is restricted to implicit learning because the monkey could solve the task without memorizing the sequence. In the present task, associative learning was imposed to assess the long-term memory more directly.

In the primate inferior temporal cortex and part of the superior temporal sulcus, neural responses to complex objects such as faces^{11,12}, hands¹² and Fourier descriptors¹³ have been reported. According to our results, acquired pairing is now regarded as an important coding faculty. The properties of pair-coding neurons indicate that memory storage is organized such that single neurons can code both pictures of the paired associates. A possible basis for this coding lies in the change of synaptic connections through repetitive learning^{14,15}, whereby two pictures are always paired with each other.

Anticipatory neural activities that precede the initiation of movements and increase during the preparatory period have been reported in the primate frontal cortex^{16–18}. In our pair-association task, the increasing delay activity of pair-recall neurons is not related to motor response because the monkey could not predict which position should be touched. As noted before, this delay activity is not only picture-selective, but also closely coupled with the paired associate that is not actually seen but retrieved. The neural mechanism for the retrieval process remains to be identified, but it may well involve the pair-recall neurons.

A recent lesion study has demonstrated that monkeys with bilateral removal of hippocampus and amygdala do not relearn the pair-association task in the training limit⁷. The type of memory this task used would therefore correspond to one that relies on the integrity of these structures. The medial temporal region is considered essential to the memory consolidation process, by which certain evanescent information obtains an enduring representation in long-term memory^{19,20}. As the anterior temporal cortex we studied links the visual system and the limbic

memory system^{21,22}, the unique neurons described here could serve as memory storage elements, also activated in the retrieval process. □

Received 14 June; accepted 24 September 1991.

- Wechsler, D. *Wechsler Memory Scale-Revised* (The Psychological Corporation, Harcourt Brace Jovanovich, San Antonio, 1987).
- Meyer, V. & Yates, A. J. *J. Neurol. Neurosurg. Psychiat.* **18**, 44–52 (1955).
- Milner, B. *Brain Mechanisms Underlying Speech and Language*, 122–145 (Grune & Stratton, New York, 1967).
- Jones, M. K. *Neuropsychologia* **12**, 21–30 (1974).
- Petrides, M. *Neuropsychologia* **23**, 601–614 (1985).
- Goldstein, L. H., Canavan, A. G. M. & Polkey, C. E. *Cortex* **24**, 41–52 (1988).
- Murray, E. A., Gaffan, D. & Mishkin, M. *Soc. Neurosci. Abstr.* **14**, 2 (1988).
- Miyashita, Y. & Chang, H. S. *Nature* **331**, 68–70 (1988).
- Miyashita, Y. *Nature* **335**, 817–820 (1988).
- Snedecor, G. W. & Cochran, W. G. *Statistical Methods* 8th edn, 97 (Iowa State University Press, Ames, 1989).
- Perrett, D. I., Rolls, E. T. & Caan, W. *Expl Brain Res.* **47**, 329–342 (1982).
- Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. J. *J. Neurosci.* **4**, 2051–2062 (1984).
- Schwartz, E. L., Desimone, R., Albright, T. D. & Gross, C. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5776–5778 (1983).
- Artola, A. & Singer, W. *Nature* **330**, 649–652 (1987).
- Frégnac, Y., Shulz, D., Thorpe, S. & Bienenstock, E. *Nature* **333**, 367–370 (1988).
- Bruce, C. J. & Goldberg, M. E. *J. Neurophysiol.* **53**, 603–635 (1985).
- Mauritz, K.-H. & Wise, S. P. *Expl Brain Res.* **61**, 229–244 (1986).
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. *J. Neurophysiol.* **61**, 331–349 (1989).
- Milner, B. *Clin. Neurosurg.* **19**, 421–446 (1972).
- Squire, L. R., Cohen, N. J. & Zola-Morgan, L. *Memory Consolidation: Psychobiology of Cognition*, 185–210 (Lawrence Erlbaum, Hillsdale, 1984).
- Insausti, R., Amaral, D. G. & Cowan, W. M. *J. comp. Neurol.* **264**, 356–395 (1987).
- Webster, M. J., Ungerleider, L. G. & Bachevalier, J. *J. Neurosci.* **11**, 1095–1116 (1991).
- Zahn, C. T. & Roskies, R. Z. *IEEE Trans. Comput.* **21**, 269–281 (1972).
- Rolls, E. T. *et al.* *J. Neurosci.* **9**, 1835–1845 (1989).
- Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. *J. Neurophysiol.* **35**, 96–111 (1972).

ACKNOWLEDGEMENTS. We thank Y. Hotta for encouragement. This work was supported by a Grant for Scientific Research on Priority Areas from the Japanese Ministry of Education, Science and Culture.

an increased risk associated with common variants at polymorphic sites in a 19-kilobase segment spanned by the 5' *INS* VNTR and the third intron of the gene for insulin-like growth factor II (*IGF2*). As *INS* is the major candidate gene from this region, diabetic and control sequences were compared to identify all *INS* polymorphisms that could contribute to disease susceptibility. In multiplex families the IDDM-associated alleles were transmitted preferentially to HLA-DR4-positive diabetic offspring from heterozygous parents. The effect was strongest in paternal meioses, suggesting a possible role for maternal imprinting. Our results strongly support the existence of a gene or genes affecting HLA-DR4 IDDM susceptibility which is located in a 19-kilobase region of *INS-IGF2*. Our results also suggest new ways to map susceptibility loci in other common diseases.

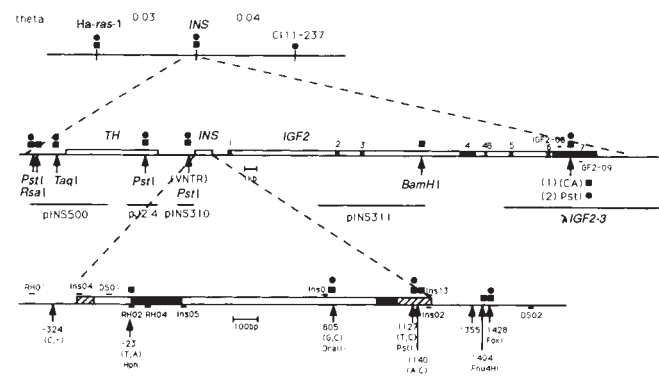


FIG. 1 Map of polymorphisms of *INS* and surrounding loci (●, characterized in isolated diabetics and controls; ■, characterized in multiplex families), and primers used for sequencing. The placement of polymorphisms is derived from Cox *et al.*²⁴, or Bell and Seino²⁵, or deduced from Southern blot hybridization of single and double digestion of DNA from homozygous individuals. *INS* coding sequences are represented in black, 5' and 3' untranslated regions by hatched lines, and introns A and B as open boxes. Primers used for PCR amplification are shown above (direct primers) and under (reverse primers) the maps. Candidate *INS* polymorphisms were determined by comparison of independent sequences from normal individuals that had been deposited in GENBANK^{9–13}, and by experiments in this study. Restriction site polymorphisms are designated by their position with respect to the first base of the initiating ATG (designated by +1), followed by the restriction enzyme; other polymorphisms are designated by the position followed by the base-pair change. Generally, for the 5' VNTR, we follow the nomenclature of Bell *et al.*¹⁴ for different allele size classes (1 represents 570-base pair (bp) mean; 2, 1,320-bp mean and 3, 2,470-bp mean). Alleles showing a smaller amount of size variation can be distinguished in each size class, but class assignment is unambiguous because of the large differences in the sizes of alleles in different classes. Presence or absence of a restriction site in haplotypes has been designated by the use of 'p' or 'a'; other polymorphisms are designated by the base-pair variant. Comparison of sequences in GENBANK led to the identification of several potential polymorphisms including four (–23/*HphI*, 805/*DraIII*, 1,127/*PstI* and 1,140 (A, C)) that were identified as base-pair changes between the two insulin alleles of one individual¹¹. New sequence data were obtained: the upstream untranslated region between the *INS* 5' VNTR and the initiating ATG (nucleotides –552 to –19) were sequenced in both haplotypes of six diabetics and eight controls; a single patient and a single control were selected for sequencing, on both haplotypes, of the remainder of *INS* (1,621 bp). After preliminary analysis, we selected a diabetic who possessed two different haplotypes associated with elevated risk of disease for further sequencing in the latter experiment. The two haplotypes contained the alleles 1paaC and 1apaA (order: VNTR) (–23/*HphI*) (805/*DraIII*) (1,127/*PstI*), (1,140 (A, C)). The control was homozygote for the haplotype associated with the lowest risk for disease, 3appA. These experiments confirmed the existence of the four potential polymorphisms identified above, and identified four new polymorphisms: at –324 (C insertion) in the 5' region, and 1,355 (C, T), 1,404/*Fnu4HI*, 1,428/*FokI* in the 3' flanking sequences. The latter three differences were also found between published sequences. Although other differences between published sequences were found, these were not confirmed by sequence data or PCR experiments done on a set of 10 unrelated individuals possessing a variety of haplotypes; therefore we conclude that they are likely to be due to sequencing errors.

Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility

C. Julier*, R. N. Hyer†, J. Davies*, F. Merlin*, P. Soularue*, L. Briant‡, G. Cathelineau§, I. Deschamps||, J. I. Rotter¶, P. Froguel*, C. Boitard||, J. I. Bell†# & G. M. Lathrop*

* Centre d'Etude du Polymorphisme Humain, 27 rue Juliette Dodu, 75010 Paris, France

† Molecular Immunology Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

‡ INSERM U100, Hôpital Purpan, Place du Docteur Baylac, 31059 Toulouse, France

§ Service d'endocrinologie, Hôpital St Louis, 2 Place du Docteur Fournier, 75010 Paris, France

|| INSERM U25, Hôpital Necker, 161 Rue de Sèvres, 75015 Paris, France

¶ Division of Medical Genetics, Cedars-Sinai Medical Center, and UCLA School of Medicine, Los Angeles, California 90048, USA

A CLASS of alleles at the VNTR (variable number of tandem repeat) locus in the 5' region of the insulin gene (*INS*) on chromosome 11p is associated with increased risk of insulin-dependent diabetes mellitus (IDDM)^{1–6}, but family studies have failed to demonstrate linkage^{5,7}. *INS* is thought to contribute to IDDM susceptibility but this view has been difficult to reconcile with the lack of linkage evidence^{6–8}. We thus investigated polymorphisms of *INS* and neighbouring loci in random diabetics, IDDM multiplex families and controls. HLA-DR4-positive diabetics showed

* To whom correspondence should be addressed.

TABLE 1 Polymorphism data

Locus	Designation of polymorphisms	Primers for PCR amplification	PCR annealing temperature	Detection method	Disease-associated or most common variant (+)
(a) Characterization of polymorphisms in the <i>INS</i> region					
<i>Ha-ras-1</i> <i>D11S454</i> <i>TH</i>	pEJ6.6/ <i>TaqI</i>				0.9 kb
	cCl11-237/ <i>TaqI</i>				1.25 kb
	pINS500/ <i>PstI</i>				2.8 kb, 2.0 kb
	pINS500/ <i>RsaI</i>				1.3 kb
	pINS500/ <i>TaqI</i>				2.2 kb
<i>IGF2</i>	pJ2.4/ <i>PstI</i>				800 bp (mean)
	pINS311/ <i>BamHI</i>				570 bp (mean)*
	lambIGF2-3/ <i>PstI</i> (CA)	igf2-08/-09	62 °C		—
<i>INS</i>	pINS310/ <i>PstI</i>			sequencing	—
	−324 (−, C)				present (p)
	−23/ <i>HphI</i>	ins04/ins05	64 °C		absent (a)
	805/ <i>DraIII</i>	ins01/ins02	56 °C		absent (a)
	1,127/ <i>PstI</i>	ins01/ins02	56 °C		C
	1,140 (A, C)	ins01/ins02	56 °C	ASO:D3564 (A, C)	C
	1,355 (C, T)	MAS (C, T)/DS02	56 °C	AS-PCR	C
	1,404/ <i>Fnu4HI</i>	ins13/DS02	62 °C		present (p)
	1,428/ <i>FokI</i>	ins13/DS02	62 °C		absent (a)
(b) Oligomers used for PCR polymorphisms typing (5' to 3')					
igf2-08	CTCATACTTTATGCATCCCCG				
igf2-09	GCCTGATCCATACAGATATCG				
ins04	TCCAGGACAGGCTGCATCAG				
ins05	AGCAATGGGCGGTTGGCTCA				
ins01	GAAGGAGGTGGGACATGT				
ins02	GCTGGTTCAAGGGCTTTA				
MAS (C, T)	CCCAAAGCGGCATGCATGT (C, T)				
ins13	TAAAGCCCTGAACCAGC				
DS02	CAGCCAGCCTCCTCCCTCCACA				

Characterization of polymorphisms in the *INS* region, and oligomers used for polymerase chain reaction (PCR) typing. *INS* polymorphisms are designated by their probe/enzyme combination or by their position (where +1 is the first base of the initiating ATG) follows the nomenclature indicated in the legend to Fig. 1. Polymorphisms at *Ha-ras-1*, *D11S454*, *TH*, *IGF2* (pINS311) and *INS* (pINS310) have been described^{26,27}. Strong linkage disequilibrium was observed between different *INS* sites. A polymorphism at 1,355 (C, T) was found to be in complete linkage disequilibrium with 1,404/*Fnu4HI* and 1,428/*FokI* in 30 individuals, and it was not characterized further. An insertion-deletion polymorphism at −324 (−, C) was present in only two out of eight normals and 0 out of six diabetics, and 1,355 (C, T), and it was not characterized in the complete series. The principal disease-associated haplotype was composed of the variants: 1paaCCpa ((5' VNTR) (−23/*HphI*) (805/*DraIII*) (1,127/*PstI*) (1,140 (A, C)) (1,355 (C, T)) (1,404/*Fnu4HI*) (1,428/*FokI*)). Members of multiplex families were characterized for polymorphisms detected with *TaqI* or *PstI*, and three PCR type polymorphisms (805/*DraIII*, 1,127/*PstI* and 1,428/*FokI*). This selection of polymorphisms allowed the parental origins of almost all haplotypes to be deduced (95% heterozygosity in the parents). Complete information could be obtained when considering the two flanking loci (*Ha-ras-1* and *Cl11-237*). Typing at probe/enzyme combination polymorphisms was done by Southern blot hybridization as described²⁸. Other polymorphisms were characterized by PCR amplification followed by specific restriction enzyme digestion (when required) and gel electrophoresis, or dot blotting and allele-specific oligomers (ASOs) hybridization, or by allele-specific amplification (AS-PCR). New polymorphisms were found in a CA-rich region in exon 7 of *IGF2*: a PCR-based assay, *IGF2*/(CA), allows detection of at least five alleles on agarose gel; probe λ *IGF2-3*/*PstI* also detects an insertion/deletion type polymorphism.

* Denoted as size class 1, following Bell *et al.*¹⁴.

Index patients from 128 multiplex families and 180 random diabetics were studied for candidate mutations in the insulin gene and surrounding loci. Polymorphisms at several adjacent loci were characterized to define the margins of the disease-associated region (Fig. 1 and Table 1). Genotype frequencies in patients and controls were not different at the *TH* or *IGF2* loci (Table 2a), nor at *Ha-ras-1* or *D11S454* (data available on request). By contrast, when alleles at the *INS* 5' VNTR were classified into three groups according to the size of the repeated region as described in Fig. 1, we found that diabetics were homozygous for the most common class (designated class 1) more frequently than controls ($P < 0.005$)¹⁻⁶.

As *INS* is the principal candidate gene in the region, all polymorphisms that could have a role in disease susceptibility in and near *INS* were identified by comparison of published sequences⁹⁻¹³ and sequences obtained for this study (Fig. 1). Of the nine *INS* polymorphisms identified (Table 1), only the 5' VNTR¹⁴ and a *PstI* variant at position 1,127 (ref. 15) have been previously characterized. Strong linkage disequilibrium was observed between the class 1 alleles of the 5' VNTR and the most frequent *INS* variants at other sites, and diabetics showed

an elevated relative risk for homozygosity of the common variants at all *INS* polymorphic sites (Table 2b). Because of the established role of HLA in IDDM susceptibility¹⁶, we also examined *INS* polymorphisms after classifying diabetics by *HLA-DR* and *DQ* genotype. The *INS* effect was restricted to the *HLA-DR4* positive diabetics (Table 2b), supporting the suggestion that there is heterogeneity in the HLA contribution to IDDM. Overall, *INS* relative risks (2.6–3.4 in all diabetics) are roughly equal to that of *DR4* itself (3.2 relative to *DRX/X* in our data). *DR4* susceptibility was seen only in association with *INS*-encoded susceptibility, demonstrating the importance of this interaction to the expression of *HLA-DR4*-mediated IDDM susceptibility. A dose effect was also evident in this *DR4* interaction, with *DR4/4* diabetics having the highest frequency of *INS* homozygosity (for the 1,127/*PstI* polymorphism in all diabetics, 31/33 *DR4/4* were *PstI* +/+ homozygotes). Subdivision of *DR4* into *DQ* subtypes *DQB1*0301* and *DQB1*0302* showed the expected increase of *DQB1*0302* in diabetic *DR4* individuals. No conclusion could be reached on *INS* susceptibility for *DQ* subtypes because too few *DQB1*0301* individuals were available for analysis.

TABLE 2 Polymorphisms of *TH*, *INS* and *IGF2* genes, and interaction between *HLA* and *INS*

(a) Counts (%) of homozygotes and nonhomozygotes in diabetics and controls

Gene	Variant	Controls (+ is disease-associated or most common variant)			Diabetics	
		++	+- or --	Series	++	+- or --
<i>TH</i>	pINS500/ <i>Rsa</i> I	32 (42%)	45 (58%)	R	83 (49%)	87 (51%)
	pINS500/ <i>Pst</i> I	28 (39%)	44 (61%)	R	86 (49%)	90 (51%)
				I	61 (51%)	58 (49%)
	pINS500/ <i>Taq</i> I	18 (23%)	61 (77%)	R	37 (22%)	135 (78%)
				I	25 (20%)	99 (80%)
<i>INS</i>	pJ2.4/ <i>Pst</i> I	20 (28%)	52 (72%)	R	51 (29%)	123 (71%)
				I	38 (32%)	80 (68%)
	5' VNTR	42 (53%)	38 (47%)	R	134 (75%)	44 (25%)
				I	90 (71%)	37 (29%)
	-23/ <i>Hph</i> I	42 (53%)	37 (47%)	R	133 (75%)	45 (25%)
	805/ <i>Dra</i> III	46 (58%)	34 (42%)	R	142 (79%)	38 (21%)
				I	87 (69%)	39 (31%)
	1,127/ <i>Pst</i> I	48 (60%)	32 (40%)	R	150 (83%)	30 (17%)
				I	94 (75%)	32 (25%)
	1,140 (A, C)	42 (53%)	37 (47%)	R	134 (75%)	44 (25%)
<i>IGF2</i>	1,404/ <i>Fnu</i> HI	46 (59%)	32 (41%)	R	150 (83%)	30 (17%)
	1,428/ <i>Fok</i> I	47 (59%)	32 (41%)	R	150 (83%)	30 (17%)
				I	93 (74%)	32 (26%)
	pINS311/ <i>Bam</i> HI	28 (36%)	50 (64%)	R	51 (37%)	87 (63%)
	<i>IGF2</i> (CA)	39 (51%)	37 (49%)	R	27 (42%)	37 (58%)

(b) Relative risks by *HLA* type for homozygote status at the *INS* loci

HLA	Number	Relative risk (95% confidence interval)			
		5' VNTR	805/ <i>Dra</i> III	1,127/ <i>Pst</i> I	1,428/ <i>Fok</i> I
All*	307	2.5 (1.5-4.1)¶	2.2 (1.3-3.7)¶	2.6 (1.6-4.4)¶	2.7 (1.6-4.5)¶
DR3/4†	91	2.6 (1.4-4.9)¶	2.1 (1.1-3.9)¶	2.9 (1.5-5.7)¶	2.8 (1.4-5.4)¶
DR4/X and DR4/4‡	104	4.2 (2.2-8.1)#	4.6 (2.3-9.3)#	5.0 (2.4-10.3)**	5.1 (2.4-10.5)**
DQB*0302§	88	4.2 (2.1-8.4)#	5.0 (2.1-10.6)#	5.6 (2.5-12.3)**	5.7 (2.6-12.7)**
DQB*0301	13	2.7 (0.8-9.2)	2.2 (0.7-7.6)	2.0 (0.6-6.9)	2.1 (2.6-7.0)
DR3/X and DR3/3	74	1.6 (0.9-3.1)	1.4 (0.8-2.8)	1.5 (0.8-3.0)	1.6 (0.8-3.0)
DRX	34	1.4 (0.7-3.2)	1.1 (0.5-2.3)	1.6 (0.7-3.6)	1.6 (0.7-3.7)

180 random IDDM patients (R), 127 index patients from multiplex families (I), and 80 controls were studied. Genotypes for 11p loci are shown by homozygote status for the most frequent allelic variant. All diabetics were ketone-positive at the time of diagnosis, and required daily insulin treatment. Random diabetics were all resident in France, and had no siblings with IDDM at the time of entry into the study. Multiplex families were collected in France, USA and North Africa. No evidence of heterogeneity of *HLA* or *INS* polymorphisms was found when families were classified by origin. Controls were parents from 40 CEPH reference families. They did not differ significantly at *INS* VNTR or *HLA* from controls reported in other studies²⁹. Contingency table analyses showed significant heterogeneity of the *HLA*-specific relative risks for all loci that were characterized in both random diabetics and multiplex probands (shown in b). *P*-values from these analyses are: 5' VNTR ($P < 0.02$), 805/*Dra*III ($P < 0.003$), 1,127/*Pst*I ($P < 0.01$), 1,428/*Fok*I ($P < 0.02$). The tests are not independent because of linkage disequilibrium. Controls showed no evidence of variation in *INS* genotype frequencies for different *HLA* genotypes; therefore, the risk estimates in Table 3 have been calculated relative to the pooled control series. Although *Ha-ras-1*, which is outside the region between *TH* and *IGF2* has been associated with IDDM in another study³⁰, the genotype frequencies at this locus do not differ in diabetics and control in our study (data not shown). *HLA*-typing was done by hybridization of complementary DNA probes DRbeta, DQbeta and DQalpha on Southern blots of *Taq*I digests as described³¹. *DR* alleles were then grouped in four categories: DR4DQB1*0302, DR4DQB1*0301, DR3, DRX (non DR3, non DR4). Contingency table analysis with log-linear models from the BMDP program package³² was applied to judge heterogeneity of the risks by *HLA* genotypes. For polymorphisms characterized both in random diabetics and in multiplex families, we did a three-way contingency table analysis to test for heterogeneity of the risks.

* Includes four individuals with unknown *HLA* genotypes.

† Includes one DR3/DR4-DQB0301 genotype.

‡ Includes three *HLA*-DR4 alleles with unknown *HLA*-DQ.

§ Includes one DR4-DQ0301/DR4-DQ0302 genotype.

¶ $P < 0.05$.¶ $P < 0.005$.# $P < 0.001$.** $P < 0.0001$.

These results led us to re-examine the question of linkage. Most previous studies have applied affected sib-pair methods to determine if the identity-by-descent (IBD) probabilities exceed 50% for *INS* haplotypes in multiplex families to evaluate the evidence for linkage^{5,7,8}. Our data suggest that many haplotypes carrying distinct class 1 alleles at the 5' VNTR loci are identical at the other *INS* sites, which are themselves likely candidates for disease susceptibility. When a parent is homozygous +/+ at candidate *INS* sites (where + represents the variant associated with IDDM), both haplotypes should be transmitted with equal frequency to affected offspring even if they can be distinguished by different class 1 alleles at the 5' VNTR. In these meioses, we expect to find no evidence of linkage. On the other

hand, when a parent is heterozygous +/- at the candidate *INS* sites (where a minus sign represents any variant other than +), the + allele should be transmitted more frequently to the affected offspring; therefore, we expect that IBD probabilities will exceed 50% in these meioses. As many meioses in multiplex IDDM families fall into the first category, where no linkage is expected, the statistical power to detect linkage will be reduced if all meioses are considered simultaneously in affected sib-pair analysis. To maximize the power to detect linkage, only the second category of meioses should be examined.

To test the effect of parental heterozygosity, polymorphisms at *INS* and neighbouring loci were used to identify four unique haplotypes in IDDM families. As expected, many parents

TABLE 3 IBD and segregation of alleles at *INS* locus

(a) Paternal and maternal IBD for diabetic offspring compared with probands in multiplex families

Variant	Alleles from father		Alleles from mother		Combined	
	1	0	1	0	1	0
Haplotype	68	54	50	72	118	126
5' VNTR	31	13†	17	15	48	28*
805/ <i>Dra</i> III	30	13†	12	14	42	27
1,127/ <i>Pst</i> I	29	10†	12	10	41	20†
1,428/ <i>Fok</i> I	30	10†	12	11	42	21†

(b) Segregation of paternal and maternal insulin alleles in relation to the HLA type of diabetic offspring in data from this study and GAW

HLA genotypes and *INS* allele transmitted to diabetic offspring from informative (+/-) meioses#

Variant study	Parental origin of <i>INS</i> allele	DRX/X		DR3/X, DR3/3		DR4/X, DR4/4		DR3/4		All DR4	
		+	-	+	-	+	-	+	-	+	-
1,127/ <i>Pst</i> I This study	father	2	3	7	11	22	8*	20	7*	42	15‡
	mother	0	0	2	7	7	6	15	14	22	20
	total	2	3	9	19	29	14*	35	21	64	35†
5' VNTR This study	father	2	3	8	13	22	9*	24	9*	46	18‡
	mother	0	0	6	8	11	9	20	17	31	26
	total	2	3	14	21	33	18*	44	26*	76	44*
GAW	father	0	0	8	6	15	4*	17	7*	32	11‡
	mother	1	2	3	3	10	4	13	10	23	14
	total	1	2	11	9	25	8†	30	17	55	25‡
Combined¶	father	2	3	15	17	36	12‡	37	15†	73	27
	mother	1	2	9	11	19	12	29	25	48	37
	total	3	5	24	28	55	24‡	66	40†	121	64§

IBD statistics were obtained by counting the number of haplotypes shared by diabetic probands and their affected sibs. Overall, 29 patients shared two haplotypes identical to those of the proband, 60 shared one, and 33 shared none. Expectations for these classes without linkage are 30.5, 61 and 30.5, respectively. In *a*, where results are presented for meioses that were informative for each of the *INS* loci, *P*-values for the test of linkage were calculated from two-sided chi-squared tests. The tests are not independent because of linkage disequilibrium. Alleles of the 5' VNTR locus have been combined into size classes as described in Table 1. *b*, The alleles that have been transmitted to the affected offspring in informative meioses, after children were classified by their HLA type. Families from the GAW study that had recombination between *INS* region markers were deemed to contain genotype errors and were removed before analysis. The two studies included eight informative families in common which were removed from the GAW data before their combination. Contingency table analysis of the 5' VNTR data from both studies showed significant heterogeneity of the transmission frequencies by HLA genotype of the affected offspring ($P < 0.025$) and by parental sex ($P < 0.04$). *P*-values were calculated from two-sided chi-squared tests for deviations from random transmission of the *INS* allele in each HLA-DR group. * $P < 0.05$; † $P < 0.01$; ‡ $P < 0.001$; § $P < 10^{-4}$; || $P < 10^{-5}$.

¶ After correction for families common to both studies.

Plus, transmission of IDDM-associated allele; minus, transmission of another allele.

carried two 5' VNTR class 1 alleles but were homozygous at other *INS* sites. As before, we found no evidence of linkage when these meioses were included in the counts of haplotype sharing (Table 3a). But in those offspring whose parents were heterozygous for one or more of the IDDM-associated *INS* variants, the IBD probabilities were significantly greater than 50% at three of the four *INS* sites studied. Preferential transmission of 5' VNTR alleles was seen in meioses from parents who carried a single class 1 allele. Surprisingly, maternal IBD probabilities did not differ from 50%, whereas the paternal effect was significant at all *INS* sites characterized in the multiplex families (Table 3a). No preferential segregation was seen in nondiabetic offspring, or in offspring of reference families obtained from the Centre d'Etude du Polymorphisme Humain (data not shown).

As the *INS* association was significant only in HLA-DR4-positive diabetics, we further subdivided meioses by the HLA genotypes of the offspring. The IDDM-associated allele *Pst*I *a* (1,127/*Pst*I) was transmitted to 38 of 50 HLA-DR4-positive diabetic offspring in informative male meioses ($P < 0.0005$), whereas 8 of 21 non-DR4 diabetics received this allele (Table 3b). The effect of maternally transmitted alleles is not significant for any genotype. Data from other loci, including the 5' VNTR with alleles grouped into size classes, showed similar segregation patterns (Table 3b). We thus decided to reanalyse data from

the 5th Genetic Analysis Workshop (GAW) which includes HLA and 5' VNTR genotypes for 94 IDDM families (data obtained from F. Clerget-Darpoux). When we classified the 5' VNTR alleles by their size, the results were remarkably similar to ours (Table 3b).

Previous studies of IDDM susceptibility have focused on insulin as a candidate gene because of its β -cell-specific expression. We have identified all mutations in or near *INS* that could account for a contribution to diabetes susceptibility. As none of the mutations alter the sequence of the protein, an *INS* effect may be due to altered regulation of gene expression. Alternatively, the *IGF2* gene, or an unidentified gene product from this region may be responsible for susceptibility.

The linkage in this study was observed principally in male meioses. Maternal-fetal interaction or genomic imprinting could account for this result. We favour the latter explanation because of the previously documented maternal imprinting of genes in this region (11p15)¹⁷⁻²⁰. *Igf2* is known to be imprinted in the mouse¹⁸, and the *INS-IGF2* synteny is conserved between species¹⁸. In man, loss of maternal imprinting of 11p15 may be implicated in the Beckwith-Wiedemann syndrome^{19,20}, which is associated with both islet cell hyperplasia and hyperinsulinaemia^{21,22}. Maternal imprinting could also account for the increased risk of disease in offspring of diabetic fathers compared with diabetic mothers²³.

Finally, our data suggest a resolution of the long-standing paradox of undetectable linkage in the presence of a strong disease association. Mapping of genetic susceptibility loci in other common diseases may also need to account for common susceptibility alleles, gene interaction and preferential effects of transmission through male or female meioses. □

Received 26 June; accepted 4 October 1991.

1. Bell, G. I., Horita, S. & Karam, J. H. *Diabetes* **33**, 176-183 (1984).
2. Bell, G. I. *et al.* *Am. J. hum. Genet.* **37**, A188 (1985).
3. Hitman, G. A. *et al.* *Diabetologia* **28**, 218-222 (1985).
4. Ferns, G. A. *et al.* *J. med. Genet.* **23**, 210-216 (1986).
5. Cox, N. J., Baker, L. & Spielman, R. S. *Am. J. hum. Genet.* **42**, 167-172 (1988).
6. Thomson, G. *et al.* *Genet. Epidemiol.* **6**, 155-160 (1989).
7. Cox, N. J. & Spielman, R. S. *Genet. Epidemiol.* **6**, 65-69 (1989).
8. Spielman, R. S. *et al.* *Genet. Epidemiol.* **6**, 43-58 (1989).
9. Bell, G. I. *et al.* *Nature* **284**, 26-32 (1980).
10. Sures, I., Goeddel, D. V., Gray, A. & Ullrich, A. *Science* **208**, 57-59 (1980).
11. Ullrich, A., Dull, T. J., Gray, A. & Sures, I. *Science* **209**, 612-615 (1980).
12. Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31-35 (1982).
13. Ullrich, A. *et al.* *Nucleic Acids Res.* **10**, 2225-2240 (1982).
14. Bell, G. I., Karam, J. H. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5759-5763 (1981).
15. Hitman, G. A., Kambo, P. K., Viswanathan, M. & Mohan, V. *J. med. Genet.* **27**, 97-100 (1990).

16. Thomson, G. *et al.* *Am. J. hum. Genet.* **43**, 799-816 (1988).
17. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849-858 (1991).
18. Beechey, C. V., Cattanch, B. M. & Searle, A. G. *Mouse Genome* **67**, 64-65 (1990).
19. Koufos, A. *et al.* *Am. J. hum. Genet.* **44**, 711-719 (1989).
20. Brown, K. W., Williams, J. C., Maitland, N. J. & Mott, M. G. *Am. J. hum. Genet.* **46**, 1000-1001 (1990).
21. Grunt, J. A. & Enriquez, A. R. *Yale J. biol. Med.* **45**, 15-21 (1972).
22. Pettenati, M. J. *et al.* *Hum. Genet.* **74**, 143-154 (1986).
23. Warram, J. H. *et al.* *New Engl. J. Med.* **311**, 149-152 (1986).
24. Cox, N. J., Bell, G. I. & Xiang, K.-S. *Am. J. hum. Genet.* **43**, 495-501 (1988).
25. Bell, G. I. & Seino, S. in *Molecular Biology of Islets of Langerhans* (ed. Okamoto, H.), (Cambridge University Press, 1990).
26. Human Gene Mapping 10. *Cytogenet. Cell Genet.* **51**, 1-1148 (1989).
27. Human Gene Mapping 10.5. *Cytogenet. Cell Genet.* **55**, 1-785 (1990).
28. Hyer, R. N. *et al.* *Am. J. hum. Genet.* **48**, 243-257 (1991).
29. Bell, G. I. *et al.* in *Molecular Approaches to Polygenic Disease*, Ciba Found. Symp. 130 (Wiley, Chichester, 1987).
30. Owerbach, D., Gunn, S. & Gabbay, K. H. *Diabetes* **39**, 1504-1509 (1990).
31. Bidwell, J. L. *et al.* *Transplantation* **45**, 640-646 (1988).
32. Dixon, W. J. *et al.* *BMDP Statistical Software Manual* (University of California Press, 1988).

ACKNOWLEDGEMENTS. We thank G. Bell, J. Mallet and T. Tokino for probes; G. Bell, H. Cann, J. Todd and A. Cambon-Thomsen for discussion; the Association Française contre les Myopathies (AFM), the Caisse Nationale de la Prévoyance and the RATP for help in collecting families; the Human Biological Data Interchange for families obtained through the Coriell Institute, and J.R., P.F., C.B., I.D., G.C. and L.B. for family collection. This work was supported by the AFM (C.J., G.M.L.), the Juvenile Diabetes Foundation (G.M.L.), the Human Frontiers Science Program (G.M.L., C.J.), and the Wellcome Trust (R.N.H., J.L.B.). J.L.B. acknowledges the Stuart Foundations and the Cedars-Sinai Board of Governors' Chair in Medical Genetics.

Deletion of steroid 5 α -reductase 2 gene in male pseudohermaphroditism

Stefan Andersson, David M. Berman,
Elizabeth P. Jenkins & David W. Russell*

Department of Molecular Genetics, The University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

THE conversion of testosterone into dihydrotestosterone by steroid 5 α -reductase is a key reaction in androgen action, and is essential both for the formation of the male phenotype during embryogenesis and for androgen-mediated growth of tissues such as the prostate^{1,2}. Single gene defects that impair this conversion lead to pseudohermaphroditism in which 46 X, Y males have male internal urogenital tracts, but female external genitalia³. We have described the isolation of a human 5 α -reductase complementary DNA from prostate⁴. Subsequent cloning and genetic studies showed that this gene (designated 5 α -reductase 1) was normal in patients with 5 α -reductase deficiency²⁶. We report here the isolation of a second 5 α -reductase cDNA by expression cloning and the polymerase chain reaction. The biochemical and pharmacological properties of this cDNA-encoded enzyme (designated 5 α -reductase 2) are consistent with it being the major isozyme in genital tissue. A deletion in this gene is present in two related individuals with male pseudohermaphroditism caused by 5 α -reductase deficiency. These results verify the existence of at least two 5 α -reductases in man and provide insight into a fundamental hormone-mediated event in male sexual differentiation.

To isolate cDNAs encoding additional 5 α -reductases, a size-fractionated and oriented cDNA library was constructed from human prostate poly(A)⁺ messenger RNA in a pCMV expression vector⁵. The size of cDNA pools that could be screened for expression of 5 α -reductase in cultured 293 cells was determined in serial dilution transfection experiments using an expression vector containing the 5 α -reductase 1 cDNA and an irrelevant cDNA library. Enzyme activity in transfected cells could be detected over background (threefold) when the 5 α -reductase 1 cDNA was diluted 10⁴-fold. On the basis of this information, 200 pools of this size from the prostate cDNA library were transfected into 293 cells and assayed for expression of 5 α -reductase activity as described in the legend to Fig. 1.

Two pools were identified which expressed a 5 α -reductase activity.

Simultaneously with the expression cloning studies, pairs of oligonucleotides derived from conserved regions of the human 5 α -reductase 1 and rat cDNAs⁶, were used in polymerase chain reactions⁷ that used cDNA template reverse transcribed from human prostate RNA. One pair, derived from the region of the cDNAs encoding the carboxy terminus of the protein, generated *inter alia* a 91-base pair product whose DNA sequence was 57% identical to the corresponding region of the human 5 α -reductase 1 cDNA. When this product was used to screen an expressing pool of prostate cDNAs, hybridization positives were obtained at a frequency of about 1 in 10⁴. This result, combined with DNA sequence analysis of a hybridization-positive clone (see below), indicated that both approaches had identified the same cDNA.

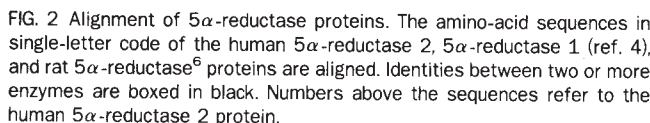
The DNA sequence of the 2.437-kilobase cDNA insert in the expression plasmid encoded a hydrophobic protein of 254 amino acids and contained a long 3'-untranslated region (Fig. 1). The sequence of the predicted protein (5 α -reductase 2) was 50% identical to that of human 5 α -reductase 1 and 46% identical to the rat 5 α -reductase enzyme (Fig. 2). All three proteins shared almost identical hydropathy plots despite their relatively low sequence identity (not shown). A search of the data bases indicated that residues 10-85 of human 5 α -reductase 2 shared a 38% sequence identity with residues 231-305 of the tobacco chloroplast NADP-ubiquinone oxidoreductase chain 5 protein⁸, and that residues 9-72 shared a 39% identity with residues 222-281 of the *pol* polypeptide of the Cas-Br-E murine leukemia virus⁹. Surprisingly, the entire 5 α -reductase 2 protein is 28% identical to residues 264-462 of the Epstein-Barr virus terminal proteins¹⁰, suggesting the latter proteins may bind steroids or NADPH.

Cell extracts from 293 cells transfected with the expression vector containing the 5 α -reductase 2 cDNA actively reduced testosterone to dihydrotestosterone, indicating that the cDNA does encode a 5 α -reductase enzyme. The data in Fig. 3a show that 5 α -reductase 2, like the major enzyme expressed in human genital skin fibroblasts¹¹ and prostate¹², has an acidic pH optimum whereas 5 α -reductase 1 has a more basic pH optimum.

The central role of 5 α -reductase in prostate development and maintenance¹³ has led to the discovery and development of powerful inhibitors of this enzyme as potential therapeutics for the treatment of prostate disorders such as benign prostatic hyperplasia^{14,15}. The 4-azasteroid finasteride (MK-906), competitively inhibits the major human prostate 5 α -reductase enzyme^{12,26}, but is a poor inhibitor of the 5 α -reductase 1

* To whom correspondence should be addressed.

Fibroblasts cultured from genital skin biopsies of 5 α -reductase-deficient subjects have reduced or abnormal acidic pH optimum enzyme activity^{3,16}. To determine if mutations in the gene encoding the 5 α -reductase 2 cDNA could be detected in these subjects, DNAs from multiple unrelated affected individuals were screened for gene rearrangements by Southern



expressing pool. The DNA sequence of one such cDNA was determined on an Applied Biosystems Model 373 fluorescence sequencer after subcloning fragments into bacteriophage M13 vectors.

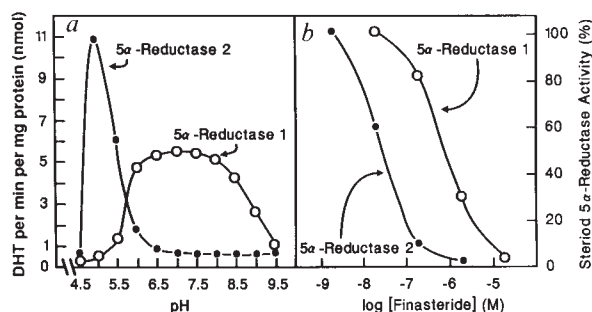


FIG. 3. Characterization of expressed 5 α -reductase isozymes. *a*, pH optima. Enzyme activity was assayed at the indicated pH in cell extracts prepared from 293 cells transfected with the 5 α -reductase 1 or 2 cDNA. Nontransfected cells express negligible amounts of enzyme activity. *b*, Inhibition by finasteride. Reductase enzyme activity obtained in extracts of transfected 293 cells in the absence of inhibitor is defined as 100%. In the presence of increasing concentrations of finasteride, progressively less 5 α -reductase enzyme activity was detected.

METHODS. An expression plasmid containing the 5 α -reductase 2 cDNA (Fig. 1) or the 5 α -reductase 1 cDNA (ref. 4) was transfected into 293 cells as described in the legend to Fig. 1. In *a*, 48 h after transfection, cell extracts were prepared as described previously⁴ and 10 μ g cell protein were assayed for 5 α -reductase enzyme activity in 0.1 M Tris-citrate buffers at the indicated pH with 10 μ M [¹⁴C]-testosterone (120 d.p.m. pmol⁻¹) as substrate and 10 mM NADPH as cofactor. In *b*, 5 μ g of transfected cell protein were assayed in duplicate for 5 α -reductase activity in the presence of the indicated concentration of finasteride (MK-906)¹² (17 β -(*N*-*t*-butyl)carbamoyl-4-aza-5 α -androst-1-ene-3-one, a kind gift of G. Rasmussen, Merck, Sharp and Dohme), 4 μ M [¹⁴C]-testosterone (120 d.p.m. pmol⁻¹) and 10 mM NADPH. In both panels, conversion into dihydrotestosterone was determined after 10-min incubations by thin layer chromatography as described previously⁶. Protein concentrations in cell extracts were measured by Lowry assay²³.

blotting. As a control, the same DNAs were screened with a probe from the 5 α -reductase 1 cDNA. The data of Fig. 4 indicate that a deletion in the 5 α -reductase 2 gene was found in two related pseudohermaphrodites from the Simbari Anga linguistic group in the Highlands of Papua New Guinea¹⁷ but was not present in the DNA of a normal individual from this tribe. The deletion seems to have removed a majority of the 5 α -reductase 2 gene from the affected individuals, as only a single weakly hybridizing fragment is visible on the autoradiogram (Fig. 4). No gross rearrangements in the 5 α -reductase 2 gene were detected in affected individuals derived from 19 different pedigrees from throughout the world, indicating that, as with many other genetic diseases¹⁸, most of the mutations cannot be detected by Southern blotting.

Our results confirm the existence of two functional 5 α -reductase genes in man. The 5 α -reductase 2 cDNA and expressed enzyme described here seems to represent the major isozyme of genital tissue based on its acidic pH optima, exquisite sensitivity to finasteride, and deletion in two related individuals with 5 α -reductase deficiency. By contrast, the 5 α -reductase 1 cDNA and enzyme is expressed at much lower amounts in the prostate⁴, has a basic pH optimum, is insensitive to finasteride, and is normal in multiple 5 α -reductase deficient patients²⁶. The isolation of two 5 α -reductases raises interesting questions for future study concerning the tissue specificity and regulation of these isozymes and their roles in androgen physiology.

Male pseudohermaphroditism reflects the incomplete establishment of phenotypic sex in 46 X, Y individuals and in a

majority of cases is due to mutations in the androgen receptor or 5 α -reductase³. These mutations are formally analogous to those described in nematodes and yeast that disrupt intercellular signalling systems controlling vulval development¹⁹ or mating type²⁷, respectively. Thus, the androgen receptor and 5 α -reductase are components of an intercellular signalling pathway that ultimately determines cell fate in the bipotential analage of the mammalian external genitalia. How the androgen receptor receives and transmits the inductive signal provided by dihydrotestosterone, as well as further parallels between mammalian and invertebrate sexual differentiation are topics for future study. □

Received 10 July; accepted 13 September 1991.

1. Wilson, J. D. *Harvey Lect.* **79**, 145-172 (1985).
2. Wilson, J. D. *Endocrinology* **25**, 491-508 (1975).
3. Griffin, J. E. & Wilson, J. D. in *The Metabolic Basis of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 1919-1944 (McGraw-Hill, New York, 1989).
4. Andersson, S. & Russell, D. W. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3640-3644 (1990).
5. Andersson, S., Davis, D., Dahlbäck, Jörnval, H. & Russell, D. W. *J. biol. Chem.* **264**, 8222-8229 (1989).
6. Andersson, S., Bishop, R. W. & Russell, D. W. *J. biol. Chem.* **264**, 16249-16255 (1989).
7. Saiki, R. K. et al. *Science* **239**, 487-491 (1988).
8. Shinozaki, K. et al. *EMBO J.* **5**, 2043-2049 (1986).
9. Rassart, E., Nelbach, L. & Jolicoeur, P. *J. Virol.* **60**, 910-919 (1986).
10. Laux, G., Perricaudet, M. & Farrell, P. J. *EMBO J.* **7**, 769-774 (1988).
11. Moore, R. J., Griffin, J. E. & Wilson, J. D. *J. biol. Chem.* **250**, 7168-7172 (1975).
12. Liang, T., Cascieri, M. A., Cheung, A. H., Reynolds, G. F. & Rasmussen, G. H. *Endocrinology* **117**, 571-579 (1985).
13. George, F. W., Russell, D. W. & Wilson, J. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8044-8047 (1991).
14. Brooks, J. R. et al. *Endocrinology* **109**, 830-836 (1981).
15. Vermeulen, A., Giagulli, V. A., DeSchepper, P., Buntinx, A. & Stoner, E. *Prostate* **14**, 45-53 (1989).
16. Moore, R. J. & Wilson, J. D. *J. biol. Chem.* **251**, 5895-5900 (1976).
17. Imperato-McGinley, J. et al. *Clin. Endocrin.* **34**, 293-298 (1991).
18. Hobbs, H. H., Russell, D. W., Brown, M. S. & Goldstein, J. L. *Rev. Genet.* **24**, 133-170 (1990).
19. Horowitz, H. R. & Sternberg, P. W. *Nature* **351**, 535-541 (1991).
20. Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).
21. Gorman, C. M., Gies, D. R. & McCray, G. *DNA Prot. Engng. Tech.* **2**, 3-10 (1990).
22. Strathmann, M., Wilkie, T. M. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7407-7409 (1989).
23. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).
24. Lehman, M. A. et al. *Science* **227**, 140-146 (1985).
25. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
26. Jenkins, E. P., Andersson, S., Imperato-McGinley, J., Wilson, J. D. & Russell, D. W. *J. clin. Invest.* (in the press).
27. Herskowitz, I. *Nature* **342**, 749-757 (1989).

ACKNOWLEDGEMENTS. We thank J. Cormier and D. Davis for technical assistance, J. McConnell for surgical samples, J. Imperato-McGinley for blood samples, J. L. Goldstein, M. S. Brown and J. D. Wilson for critical reading of the manuscript. This work was supported by the NIH, the Welch Foundation, and the Perot Family Foundation. The sequence data presented in Fig. 1 will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number M74047.

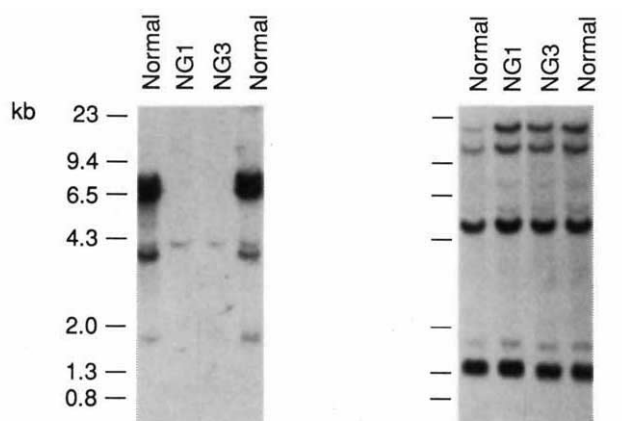


FIG. 4 Deletion of 5 α -reductase 2 gene in subjects with 5 α -reductase deficiency. DNA isolated from two normal individuals and two related 5 α -reductase deficiency subjects from a geographically isolated tribe in the Highlands of Papua New Guinea¹⁷ was screened by Southern blotting using the indicated 5 α -reductase cDNA probes. The normal DNA analysed in the most left-hand lane was isolated from an individual from the same New Guinea tribe as the NG1 and NG3 subjects, whereas the normal DNA analysed in the right-hand lane was isolated from a caucasian American. The filter was screened with the 5 α -reductase 2 cDNA probe first, and then stripped and reprobed with the 5 α -reductase 1 cDNA probe. A deletion of all but a weakly hybridizing fragment of about 4.5 kilobases in the DNA of the affected NG1 and NG3 individuals is apparent from the autoradiogram obtained with the 5 α -reductase 2 probe. All individuals have a normal 5 α -reductase 1 gene.

METHODS. Genomic DNA was isolated from peripheral blood samples from the indicated individuals and prepared and analysed by high-stringency Southern blotting as described previously²⁴. Aliquots of DNA (20 μ g) were digested with *Hind*III before electrophoresis on agarose gels. Three single-stranded ³²P-labelled probes spanning the coding region of the 5 α -reductase 2 cDNA shown in Fig. 1 were prepared as described by Church and Gilbert²⁵ and used in hybridization. After autoradiography for 5 days at -70 °C, the filter was stripped²⁰ and reprobed with a random hexanucleotide ³²P-labelled probe²⁰ corresponding to the full-length 5 α -reductase 1 cDNA (ref. 4).

FtsZ ring structure associated with division in *Escherichia coli*

Erfei Bi & Joe Lutkenhaus

Department of Microbiology, Molecular Genetics & Immunology, University of Kansas Medical Center, Kansas City, Kansas 66103, USA

GENES for cell division have been identified in *Escherichia coli* by the isolation of conditional lethal mutations that block cell division, but do not affect DNA replication or segregation¹. Of these genes, *ftsZ* is of great interest as it acts earliest in the division pathway^{2,3}, is essential⁴, its level dictates the frequency of division^{5,6}, and it is thought to be the target of two cell-division inhibitors⁷⁻⁹, Sula, produced in response to DNA damage¹⁰, and MinCD, which prevents division at old sites¹¹. Here we have used immunoelectronmicroscopy to localize the FtsZ protein to the division site. The results suggest that FtsZ self-assembles into a ring structure at the future division site and may function as a cytoskeletal element. The formation of this ring may be the point at which division is regulated.

To obtain information about the role of FtsZ in cell division, attempts were made to localize FtsZ after cell fractionation.

After cell breakage with a French press most FtsZ was localized to the cytoplasmic fraction, with only a small amount (<5%) present in the membrane fraction even if Mg^{2+} was present. Thus, if FtsZ acted at the membrane the association was weak and did not survive cell breakage. In another approach to localize FtsZ we used immunoelectronmicroscopy. Exponentially growing cells of *E. coli* MC4100 (ref. 12) were fixed with 2% glutaraldehyde and processed for immunolabelling (Fig. 1 legend). Thin sections were incubated with affinity-purified FtsZ antibodies and then incubated with colloidal gold-labelled secondary antibody. Figure 1 shows representative labelled cells at different stages of the cell cycle. The most noticeable feature of the distribution of label was a distinct pattern in dividing cells. In longitudinal sections all cells that contained a constriction contained a cluster of label at the leading edge of the invagination. This symmetrical pattern of labelling indicated that FtsZ must form a ring-like structure.

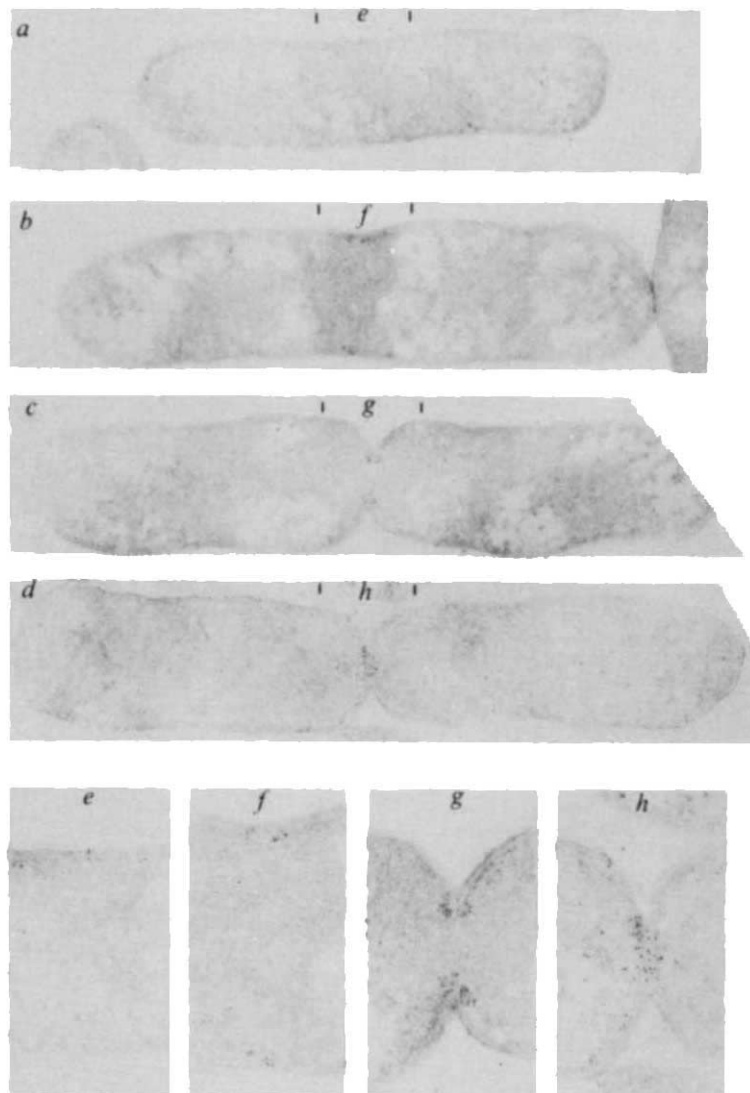
The cells in Fig. 1 are arranged to reflect progression through the cell division cycle. Examination of short cells with no sign of constriction and the DNA not obviously segregated revealed no specific pattern of labelling. The label was randomly distributed in the cytoplasm (Fig. 1a). The next class, as typified by the cell in Fig. 1b, contained cells in which the DNA was

segregated, but there was no clear sign of constriction. Such cells contained some clustering of the gold label at the cytoplasmic membrane near the longitudinal midpoint of the cell, where the next division would occur. The next class, Fig. 1c, contains cells with a clear sign of constriction. These cells showed clusters of gold label at the leading edges of the invagination. The final class of cells, Fig. 1d, contains cells that have almost completed septation but not yet separated. These cells also show clustering of label at the division site. This is the first report of a protein specifically localized to the septal region.

The gold labelling pattern in longitudinal sections suggested that FtsZ formed a ring-like structure at the division site which decreased in diameter as septation ensued. Random screening of cross-sections of cells did not reveal any particular labelling pattern. But the frequency of observing a ring structure in cross-sections would be rather low as the septum represents a small fraction of the total cell length and only about 20% of the cells are dividing. To facilitate searching for a ring structure we screened cross-sections of cells for those that lacked DNA and had a smaller diameter, two characteristics expected of septal cross-sections (also of polar regions). This step greatly aided the screening process and allowed large numbers of cross-sections to be screened rapidly. Figure 2 shows two pairs of

FIG. 1 Distribution of the FtsZ protein throughout the cell-division cycle as revealed by immunogold labelling. *a, b, c, d*, The gold-labelling pattern of cells at different stages of the cell cycle. *e, f, g, h*, Enlarged version of the middle region of the cells shown in *a, b, c* and *d*, respectively.

METHODS. Affinity purification of the FtsZ antibody: a rabbit anti-FtsZ serum was passed through an FtsZ-Sepharose 4B affinity column. Antibodies against FtsZ were eluted with 0.2M glycine, pH 2.6 and concentrated with a microconcentrator (Centricon-30, AMICON, USA). This affinity-purified antibody was used in the following experiments for probing FtsZ protein. Immunoelectronmicroscopy: the *E. coli* strain, MC4100, was grown in L-Broth (10 g tryptone, 5 g yeast extract, 5 g NaCl and 50 mg thymine per distilled water) at 37 °C. After the culture had been growing exponentially for several generations glutaraldehyde (product of Electron Microscopy Sciences) was added to the growth medium to a final concentration of 2%. Cells were fixed at 37 °C for 30 min and then collected by centrifugation at room temperature. The cell pellet was washed with 1 × PBS buffer (0.2 g KCl, 0.2 g KH_2PO_4 , 8 g NaCl, 2.16 g $Na_2HPO_4 \cdot 7H_2O$ per litre of distilled water, pH 7.3) for 10 min and then immersed in 30% ethanol for 17 min. Cells were stained with 2% uranyl acetate (dissolved in 50% ethanol) for 20 min and dehydrated in a gradient of ethanol: 70%, 80%, 95% and 100%. Each step was for 5 min, except for the dehydration with 100% ethanol which was for 10 min and was repeated once. After dehydration, the cell pellet was infiltrated in LR White resin (London Resin Company) overnight at 4 °C. Fresh resin was added to a gelatin capsule with the cell pellet at the bottom for embedment. The cells were embedded at 60 °C overnight. Thin sections were obtained by cutting the specimen on an LKB-Ultratome (NOVA). Gold- and silver-coloured sections were picked onto nickel grids for immunogold labelling. Immunogold labelling was done as follows: thin sections on nickel grids were rinsed in 0.2M Tris-HCl, pH 7.6, for 5 min, then blocked in 50 mM Tris-HCl, 0.15M NaCl and 1% BSA (TSB buffer) for 15 min. Subsequently, the grids were incubated in the affinity purified FtsZ antibody solution (in TSB buffer at a concentration of $7.4 \mu g ml^{-1}$) for 1.5 h at room temperature, followed by washing with TSB buffer twice, 5 min each time. The grids were transferred into a solution containing 5 nM gold-labelled goat anti-rabbit IgG(H+L) (Amersham; diluted 1:10 in TSB buffer) and incubated for 1.5 h at room temperature. Following secondary antibody treatment, the grids were washed with TSB buffer twice and with filtered distilled water twice, 5 min each time. Then the grids were stained with filtered aqueous 7% uranyl acetate solution for 2 min. The grids were dried and examined with an electron microscope (JEOL 100-CXII). Cellular images resulting from cutting at different angles were recorded by photography.



cells that resulted from such a screen. The smaller-diameter cells which are lacking DNA in the cross-section are labelled with gold at their periphery, confirming a ring-like structure. By contrast, cross-sections of cells containing DNA contain relatively little label which is randomly dispersed. In addition, the inner membrane and cell wall are less well defined in labelled cells than in unlabelled cells. This is as expected for cross-sections through the septal region where the thickness of the section is greater than the width of the septum and the curvature of the cell away from the plane of the septum would result in a periphery that is less dense.

The distribution of the gold label suggests a model for the role of FtsZ in cell division. In this model (Fig. 3) we propose that FtsZ self-assembles into a ring structure on the cytoplasmic surface of the inner membrane where septation will occur and that this is the first step in the division process. We also suggest that the initiation of the assembly of FtsZ into a ring structure is the rate-limiting step for septation and occurs when the amount of FtsZ is sufficient. This ring structure must dictate the septation site and act to localize septal growth perhaps by interacting directly with septal-specific peptidoglycan biosynthetic machinery at the leading edge of the invagination¹³. Thus, formation of the ring would be the key point at which temporal and spatial control over division are exerted.

Consistent with this model are the following: (1) the localization of FtsZ in dividing cells versus nondividing cells; (2) the amount of FtsZ, about 20,000 monomers per rapidly growing cell (E.B. and J.L., unpublished results), is sufficient to form a ring structure through self assembly; (3) genetic studies indicating an interaction between *fisZ* alleles demonstrating that FtsZ acts as a multimer^{7,14}; and (4) studies altering the amount of FtsZ which demonstrate that FtsZ is both essential and rate limiting for division.

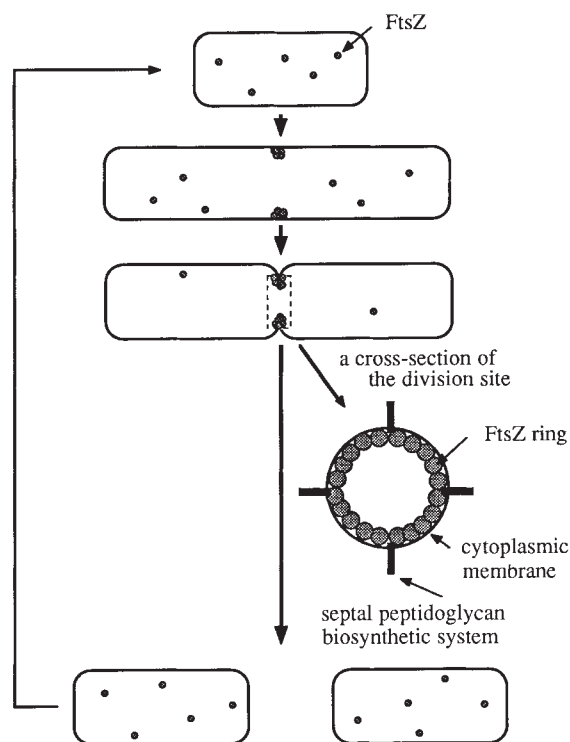


FIG. 3 Model for the role of FtsZ in cell division. See text for details.

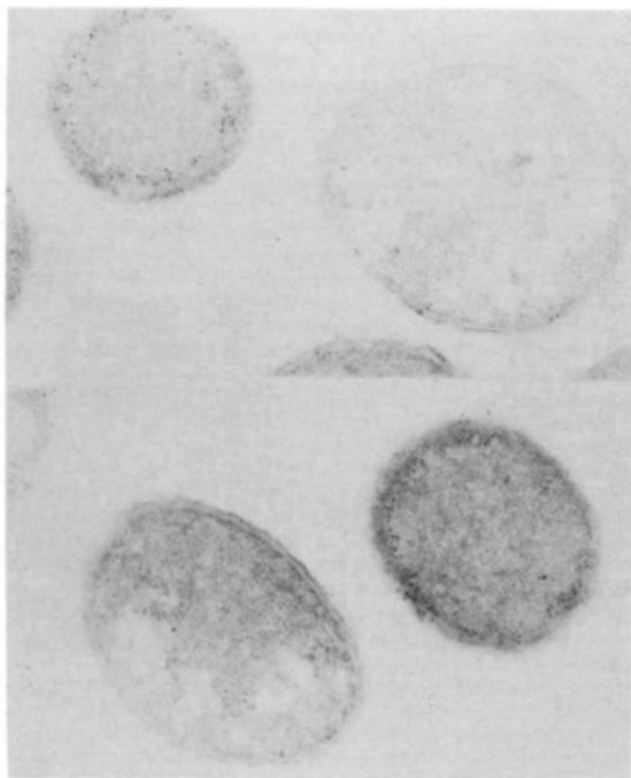


FIG. 2 Location of FtsZ near the cytoplasmic membrane in cross-sections through the septum. The methods are the same as in Fig. 1.

How the ring is localized is not known. It is possible that the ring forms at a site that exists on the membrane or forms in response to some cytoplasmic element. The existence of sites was suggested in a model to explain the phenotype of the *min* mutant. It was proposed that the cell poles are old sites that can compete with the medial site if not masked by the *min* system¹⁵. The other possibility would not require a unique site on the membrane for the structure to form, but that the formation of the structure responds to some cytoplasmic element such as the DNA segregation machinery¹⁶. Such a coupling would explain the negative effect that blocking DNA replication and/or segregation has on cell division in the absence of the SOS DNA damage-inducible response^{17,18}. It is also clear that the ring structure is dynamic and contracts during the division process. How this occurs is not clear but it may involve the participation of energy-consuming proteins such as the possible myosin-like protein¹⁹. Mutations in other known *fis* genes block cell division at a later step morphologically^{2,3} and must arrest division after formation of the ring. By contrast, the inhibitors Sula and MinCD block cell division at an early step morphologically and may prevent ring formation.

We have previously shown that FtsZ is widespread among the eubacteria²⁰ and is essential for vegetative and sporulation septation in the Gram-positive bacterium *Bacillus subtilis*²¹. Thus, the ring structure we have observed for *E. coli* may be a conserved mechanism by which eubacteria divide. Although FtsZ has no homology to any cytoskeletal proteins, in our model the role of FtsZ is to form a cytoskeletal element that is functionally analogous to the role of actin in cytokinesis in animal cells²².

The FtsZ ring structure observed in this study raises many questions, for example, what is the cell cycle signal to initiate formation of the structure as FtsZ is synthesized continuously during the cell cycle²³, what are the geometry and dynamics of the structure, and with what cellular components does the ring

interact? Further work on FtsZ should answer these questions and lead to an understanding of bacterial cell division. □

Received 1 August; accepted 16 September 1991.

1. Bi, E. & Lutkenhaus, J. in *Prokaryotic Structure and Function: A Perspective* (eds Mohan, S. B., Dow, C. & Cole, J. A.) (Cambridge University Press, Cambridge, in the press).
2. Begg, K. J. & Donachie, W. D. *J. Bact.* **163**, 615–622 (1985).
3. Taschner, P. E. M., Huls, P. G., Pas, E. & Woldringh, C. L. *J. Bact.* **170**, 1533–1540 (1988).
4. Ward, J. E. Jr & Lutkenhaus, J. F. *Cell* **42**, 399–412 (1985).
5. Bi, E. & Lutkenhaus, J. *J. Bact.* **172**, 2765–2768 (1990).
6. Dai, K. & Lutkenhaus, J. *J. Bact.* **172**, 3500–3506 (1991).
7. Bi, E. & Lutkenhaus, J. *J. Bact.* **172**, 5602–5609 (1990).
8. Bi, E. & Lutkenhaus, J. *J. Bact.* **172**, 5610–5616 (1990).
9. de Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1129–1133 (1990).
10. Huisman, O. & D'Ari, R. *Nature* **290**, 797–799 (1981).
11. de Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. *Cell* **56**, 641–649 (1989).
12. Ito, K., Bassford, P. J. Jr & Beckwith, J. *Cell* **24**, 707–717 (1981).
13. Wientjes, F. B. & Nanninga, N. *J. Bact.* **171**, 3412–3419 (1989).
14. Ward, J. E. Jr & Lutkenhaus, J. F. *J. Bact.* **157**, 815–820 (1984).
15. Teather, R. M., Collins, J. F. & Donachie, W. D. *J. Bact.* **118**, 407–413 (1974).
16. Niki, H., Jaffe, A., Imamura, R., Ogura, T. & Hiraga, S. *EMBO J.* **10**, 183–193 (1991).
17. Burton, P. & Holland, I. B. *Molec. Gen. Genet.* **190**, 309–314 (1983).
18. Jaffe, A., D'Ari, R. & Norris, V. *J. Bact.* **165**, 66–71 (1986).
19. Casaregola, S., Norris, V., Goldberg, M. & Holland, I. B. *Molec. Microbiol.* **4**, 505–511 (1990).
20. Corton, J. C., Ward, J. E. Jr & Lutkenhaus, J. *J. Bact.* **169**, 1–7.
21. Beall, B. & Lutkenhaus, J. *Genes Dev.* **5**, 447–455 (1991).
22. Schroeder, T. E. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1688–1692 (1973).
23. Robin, A., Joseleau-Petit, D. & D'Ari, R. *J. Bact.* **172**, 1392–1399 (1990).

ACKNOWLEDGEMENTS. We thank B. R. Fegley for assistance with electron microscopy. This work was supported by the NIH.

Integration of a group I intron into a ribosomal RNA sequence promoted by a tyrosyl-tRNA synthetase

Georg Mohr & Alan M. Lambowitz*

Departments of Molecular Genetics and Biochemistry and the Biotechnology Center, The Ohio State University, 484 West Twelfth Avenue, Columbus, Ohio 43210, USA

GROUP I and II introns are mobile elements that propagate by insertion into different genes¹. Some introns of both types self-splice *in vitro* by transesterification reactions catalysed by the intron RNA². These transesterifications are reversible^{3–5}, and it has been suggested that reverse splicing followed by reverse transcription and recombination with genomic DNA may be a mechanism for intron transposition^{6,7}. *In vivo* the splicing of many, if not all, group I and II introns requires protein factors, which may facilitate correct folding of the intron RNAs⁸. Here we show that the *Neurospora* mitochondrial large rRNA intron, a group I intron that is not self-splicing *in vitro*⁹, undergoes reverse splicing in a reaction promoted by the CYT-18 protein, the *Neurospora* mitochondrial tyrosyl-tRNA synthetase, which is required for splicing the intron *in vivo*¹⁰. In contrast to known RNA-catalysed reverse splicing reactions, this protein-assisted reverse splicing is sufficiently rapid to compete with forward splicing at low RNA concentrations under physiologically relevant conditions, including high GTP and low Mg²⁺ concentrations. Our results indicate that proteins that promote splicing could contribute to intron mobility by promoting reverse splicing *in vivo*.

The *Neurospora* mitochondrial tyrosyl-tRNA synthetase, which is encoded by the nuclear gene *cyt-18*, is required for splicing the mitochondrial large rRNA intron and other group I introns in *Neurospora* mitochondria¹⁰. Splicing proceeds by the same guanosine-initiated transesterification mechanism used by self-splicing group I introns, and the CYT-18 protein is

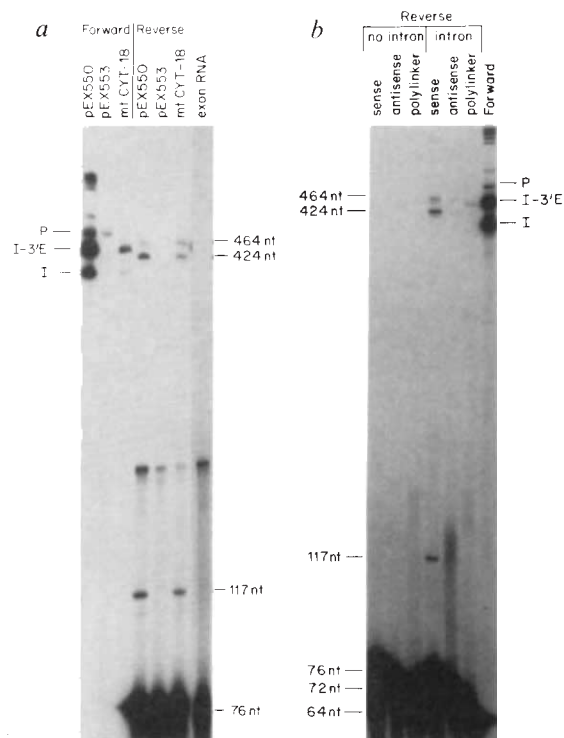


FIG. 1 Forward and reverse splicing of the *N. crassa* mitochondrial (mt) large rRNA intron. **a**, Forward and reverse splicing reactions with wild-type CYT-18 protein expressed in *E. coli* (pEX550) or isolated from mitochondria (mt CYT-18) or with an *E. coli*-synthesized mutant CYT-18 protein having an internal deletion of 72 amino acids (pEX553). Splicing was assayed using a 503-nucleotide (nt) *in vitro* transcript containing a 388-nt derivative of the mt large rRNA intron⁹ and is indicated by [³²P]-GTP-labelled bands of 389 nt (intron, I) and 439 nt (intron-3' exon, I-3'E). CYT-18 protein preparations from *E. coli* also contain a putative end-labelling activity, which labelled the 503-nt precursor (P) RNA¹³. Weak bands visible in the pEX553 reaction probably result from degradation of this [³²P]-labelled precursor RNA, rather than from splicing. Reverse splicing was assayed using the 389-nt excised intron RNA and an [³²P]-UTP-labelled, 76-nt *in vitro* transcript that contains the ligated exon sequence (Fig. 2). The lane labelled 'exon RNA' shows the 76-nt exon RNA before reaction. The asterisk denotes an additional band in the exon RNA preparation. **b**, Reverse splicing with different exon RNAs. Sense, 76-nt transcript containing the ligated exon sequence; antisense, 72-nt transcript from the strand opposite that giving the ligated exon sequence; polylinker, 64-nt polylinker transcript without the ligated exon sequence.

METHODS. Splicing reactions were as described^{9,13} in 20 µl containing 0.3 µM precursor RNA (503-nt *in vitro* transcript from pBD5A linearized with *Ban*I), 0.6 µM [³²P] GTP (3,000 Ci mmol⁻¹; New England Nuclear), 20 units RNasin (Promega), 100 mM KCl, 5 mM MgCl₂, 20 mM Tris-HCl, pH 7.5, 5 mM DTT, and CYT-18 protein preparation (pEX550, 220 nM; pEX553, 270 nM; mt CYT-18, 195 nM; CYT-18 concentrations used in assays calculated for monomer protein). Products were analysed by electrophoresis in a 4% polyacrylamide-8 M urea gel, followed by autoradiography. For reverse splicing, 4 nM each of intron RNA and [³²P]UTP-labelled 76-nt exon RNA were mixed in 20 µl of splicing buffer (see above and ref. 13) containing 30 mM MgCl₂ and incubated on ice for 5–10 min. Reactions were initiated by addition of CYT-18 protein preparations at the same concentrations used for forward splicing (see above) and incubated for 15 min at 37 °C. Further processing of samples was as described for forward reactions¹³. Radioactivity of reverse-spliced RNAs was determined using a Betascope 603 (Betagen). The intron RNA for reverse splicing was purified from forward splicing reactions similar to those in **a** by electrophoresis in 4% polyacrylamide-8 M urea gels. The 5'-terminal ³²P was removed by treatment with alkaline phosphatase, and the RNA was phenol-extracted and ethanol-precipitated. The 76-nt exon RNA was synthesized from *Eco*RI-digested pBS100, as described in Fig. 2.

* To whom correspondence should be addressed.

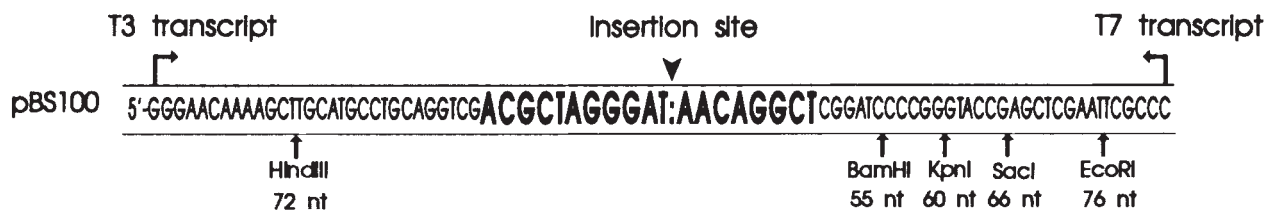


FIG. 2 Region of pBS100 used to synthesize ligated exon RNAs for reverse splicing. pBS100 was constructed from pBS(-) (Stratagene) by inserting the complementary oligonucleotides 5'-TCGACGCTAGGGATAACAGGCTCG and 5'-GATCCGAGCTGTTATCCCTAGCG between the *Sall* and *Bam*HI sites. The 76-nt exon RNA was synthesized from *Eco*RI-digested pBS100, using bacteriophage T3 RNA polymerase (Stratagene) with 23 μ M [α - 32 P]UTP (495 Ci mmol $^{-1}$; New England Nuclear). RNA was purified by electrophoresis

in 6 or 8% polyacrylamide-8 M urea gels, phenol-extracted and ethanol-precipitated. Restriction sites shown to the right were used to synthesize RNAs that differ in the length of the 3' exon. The *Hind*III site shown to the left was used to linearize pBS100 for synthesis of the 72-nt antisense transcript, using bacteriophage T7 RNA polymerase. Bold type indicates the ligated exon sequence from the *Neurospora* mt large rRNA.

presumed to function by facilitating correct folding of the precursor RNA¹¹. Figure 1a illustrates the CYT-18 protein-dependent splicing of the *Neurospora* mitochondrial large rRNA intron, using purified preparations of the CYT-18 protein isolated from mitochondria or synthesized by expression plasmids in *Escherichia coli*^{12,13}. Splicing was assayed using a 388-nucleotide derivative of the mitochondrial large rRNA intron that shows the same CYT-18 protein dependence as the full-length intron⁹. For *in vitro* forward splicing reactions, a 503-nucleotide *in vitro* transcript containing the 388-nucleotide intron was incubated with CYT-18 protein and [α - 32 P]GTP, and splicing was assayed by gel electrophoresis and autoradiography. The reaction resulted in [32 P]GTP-labelled RNAs of 389 nucleotides, corresponding to excised intron, and 439 nucleotides, corresponding to intron-3' exon intermediate. Both *E. coli*-synthesized CYT-18 protein (pEX550) and mitochondrial CYT-18 protein functioned equally well in splicing, whereas a mutant CYT-18 protein (pEX553) with an internal deletion of 72 amino acids did not support efficient splicing.

To determine whether the intron could undergo protein-dependent reverse splicing, the 389-nucleotide intron RNA generated during forward splicing was incubated with CYT-18 protein and an [α - 32 P]UTP-labelled, 76-nucleotide *in vitro* transcript containing the ligated exon sequence (Figs 1a and 2). The reaction resulted in labelled RNAs of 464 nucleotides, the size expected for fully reverse-spliced product, and 424 nucleotides, the size expected for the intron-3' exon intermediate, as well as some additional lighter bands, which may reflect integration at cryptic sites (see below). Both mitochondrial and *E. coli* synthesized CYT-18 protein (pEX550) catalysed reverse splicing, whereas no products were observed without protein or with the mutant pEX553 protein (Fig. 1a). As well as the putative reverse-spliced products, we also detected a 117-nucleotide band identified by RNase H digestion as the exon RNA (76 nucleotides) with an additional 5' exon (41 nucleotides) ligated at or near its 5' end (not shown). This molecule may be formed by ligation of a free 5' exon near the 5' end of the pBS100 transcript at the sequence G-AACA, which resembles the 3' intron-exon

FIG. 3 Reverse splicing results in precise insertion of intron RNA into ligated exons. Sequence analysis of 5' exon-intron boundary (left) and intron-3' exon boundary (right) of cDNA clone 3-1 derived from the 464-nt fully reverse spliced RNA. The sequences of the *N. crassa* mt large rRNA intron and parts of the exons are shown. A schematic of the sequencing strategy is shown below.

METHODS. For cDNA cloning of reverse spliced products, reverse splicing was done as in Fig. 1, but scaled up 10–20-fold. The 464-nt RNA was electroeluted from gel slices, and cDNA was synthesized with a BRL super-script kit (GIBCO BRL) according to the manufacturer's instructions, using 42 nM of 3' exon primer (5'-GAAGATCTGAATTCGAGCTCGGTACC) complementary to the polylinker sequence of the 3' exon derived from pBS100. The cDNA was then amplified with AmpliTaq DNA polymerase (5 units; United States Biochemical) with a 5' exon primer (5'-CAAAAGCTTGCATGCGCTGAG) and the 3' exon primer in a DNA Thermal Cycler (Perkin-Elmer Cetus) (1 min, 94 °C; 2 min 50 °C; 1.5 min 72 °C; 25 cycles). After the first round of amplification, the amplified DNA was purified by agarose gel electrophoresis, electroeluted, and used as template for a second PCR amplification. The final PCR product was digested with *Hind*III and *Bgl*II and cloned in pBS(-). White or pale blue colonies were selected and plasmid DNA isolated. Clones containing inserts of around intron size were sequenced as shown in the figure with oligonucleotide RA1 (5'-ATTTAAGGATGCGATGCTGAGGCTTACC) and forward M13 primer.

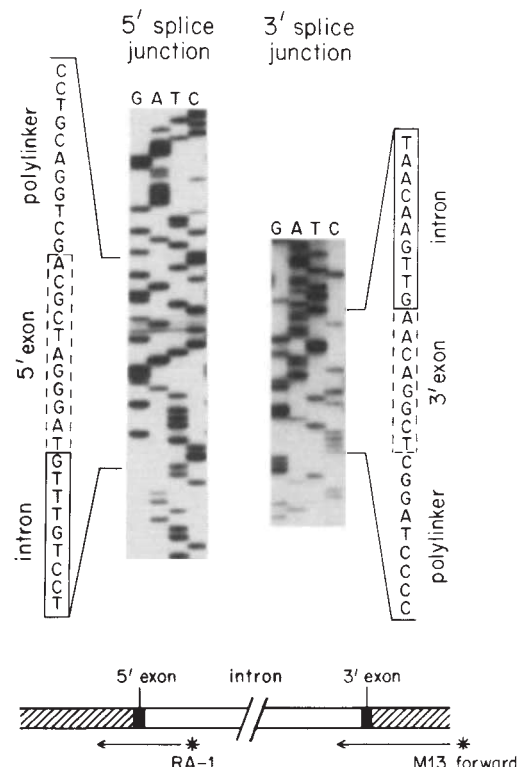
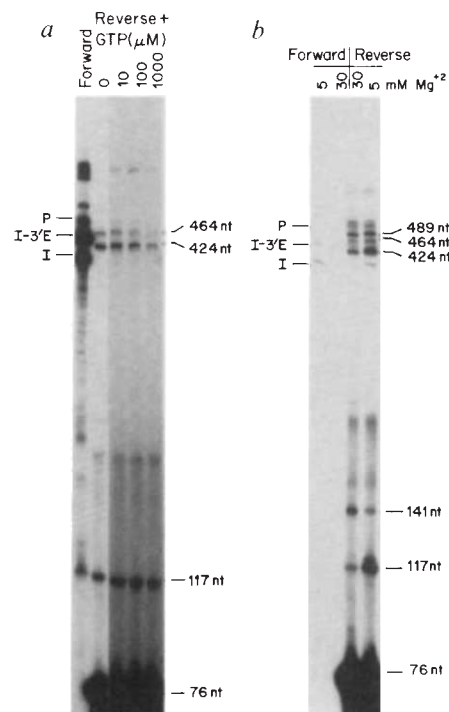


FIG. 4 Reverse splicing competes with forward splicing at high GTP and low Mg^{2+} concentrations. *a*, Reverse splicing in the presence of increasing amounts of GTP. Reverse splicing reactions contained 480 nM CYT-18 (pEX550) protein, 4 nM 389-nt intron RNA, 4 nM 76-nt exon RNA and GTP at concentrations indicated in the figure. Reactions were incubated at 42 °C for 30 min and analysed as described in Fig. 1. *b*, Forward and reverse splicing occur simultaneously under forward splicing conditions. Lanes 1 and 2 show forward splicing reactions with 0.3 μ M pED5A precursor RNA, 365 nM CYT-18 (pEX550) protein, 100 μ M [α - 32 P]GTP (10 Ci mmol $^{-1}$) and 5 or 30 mM $MgCl_2$. Incubation was at 37 °C for 1 h. Lanes 3 and 4 show identical forward splicing reactions in the presence of 4.9 nM 76-nt ligated exon RNA labelled during transcription with [α - 32 P]UTP. The 424- and 464-nt reverse spliced products were identified by comigration with authentic products in an adjacent lane (not shown). The product bands also included putative recombined RNAs, which may contain the 5' exon from pBD5A (for example, bands above 464 nt and at 141 nt). The 489-nt band may be a recombined RNA containing the 5' exon from pBD5A and the 3' exon from the 76-nt RNA. This product could be formed either by reverse splicing with exchange of 5' exons after the first step or by hydrolysis of the pBD5A precursor at the 3' splice site followed by attack at the splice junction of the 76-nt transcript. The 141-nt RNA, observed only in this type of experiment, has a size consistent with its being the 76-nt exon RNA containing an additional 65-nt 5' exon from the pBD5A precursor. The amounts of reverse-spliced products were decreased at 30 mM Mg^{2+} owing to the slower rate of forward splicing and the consequent decrease in amount of free intron RNA.



junction (compare ref. 14). Reverse splicing was also observed using a larger (583-nucleotide) derivative of the large rRNA intron, and we confirmed that the putative fully reverse-spliced product and intron-3' exon intermediate were labelled appropriately with either 5'- or 3'-end labelled exon RNAs (not shown).

Efficient reverse splicing was dependent on the large rRNA exon sequence in the target RNA. Figure 1b shows that a 64-nucleotide polylinker transcript lacking the large rRNA sequence and a 72-nucleotide antisense exon RNA gave only weak signals, presumably owing to the use of cryptic insertion sites resembling the rRNA exon sequence. Previous studies of reverse splicing of group I and II introns have shown that the intron insertion site needs only contain a short sequence related to the exon sequence normally flanking the 5' splice site³⁻⁵.

Reverse splicing of the *Neurospora* mitochondrial large rRNA intron also depended on the size of the exon RNA. Increasing the length of the 3' exon from 35 to 103 nucleotides, or shortening it to 25, 19 or 14 nucleotides, decreased the amount of reverse-spliced product up to 100-fold (not shown). The effect of exon sequences could reflect differences in secondary structure, as described for reverse splicing of the *Tetrahymena* intron¹⁵, and/or additional interactions between the exon RNAs and the intron or CYT-18 protein.

To confirm the accuracy of reverse splicing, cDNAs corresponding to the 464-nucleotide fully reverse-spliced product were cloned by polymerase chain reaction (PCR) procedures. Precursor RNAs transcribed from four of five clones tested were spliced as efficiently as the wild-type intron (not shown). The sequences of two clones that spliced efficiently showed that the intron had inserted at the correct position (Fig. 3).

Reverse splicing assayed with the 388-nucleotide intron has optima of 37 to 42 °C, 30 mM Mg^{2+} , and 25 mM KCl or less, compared to optima of 45 °C, 5 mM Mg^{2+} , and 100 mM KCl for forward splicing⁹. Under optimal conditions with 1–20 nM of exon and intron RNA, the reverse-spliced products increased linearly for around 1 h and plateaued at 5–8% of the input RNA. The rate and extent of reactions with nanomolar concentrations of exon and intron RNAs were similar to those at micromolar RNA concentrations for the *Tetrahymena* intron and group II

introns³⁻⁵, possibly reflecting the acceleration by the protein relative to the strictly RNA-mediated reactions.

In principle, protein-dependent reverse splicing could proceed in several ways. The protein may bind to the intron and ligated exons separately; it may bind to an intron-exon complex; or it may bind the intron forming an active complex, which then binds the exons. Nitrocellulose filter binding assays showed that the CYT-18 protein bound to 32 P-labelled *Neurospora* mitochondrial large rRNA intron with K_d of ~10 nM (binding assays carried out at CYT-18 protein excess and assuming the CYT-18 protein binds as a dimer; Q. Guo and A.L., unpublished data). However, the CYT-18 protein did not bind strongly to either 32 P-labelled exon RNA or 32 P-labelled exon RNA which had been preincubated with intron RNA (not shown). These results are consistent with reverse splicing proceeding by the binding of CYT-18 protein to the intron, followed by binding of the exons to an active ribonucleoprotein complex. At high concentrations of CYT-18 protein (>100 nM), we observed a parallel increase in binding of both exon RNA and a polylinker transcript lacking the exon sequences. Such nonspecific binding of the exon or polylinker RNA to the CYT-18 protein may contribute to reverse splicing *in vitro*. But in reverse splicing reactions carried out with 2 nM intron RNA, 170 nM CYT-18 protein and increasing amounts of exon RNA, the K_m value for exon RNA was 60 nM, suggesting that the affinity of the exon RNA for the intron-protein complex is substantially higher than its affinity for the protein.

Woodson and Cech found that reverse-spliced products of the *Tetrahymena* intron disappear upon addition of 25 μ M GTP, and they anticipated that reverse splicing would be even more inhibited at higher GTP concentrations *in vivo*³. By contrast, Fig. 4a shows that reverse splicing of the *Neurospora* intron occurs in the presence of GTP concentrations as high as 1 mM. Figure 4b shows that addition of 4.9 nM 32 P-labelled exon RNA to a forward splicing reaction with 5 mM Mg^{2+} and 100 μ M GTP resulted in appearance of reverse spliced products, indicating that reverse splicing competes with forward splicing, even under optimal conditions for forward splicing. The product bands included not only those expected for the synthetic exon

RNA (424 and 464 nucleotides), but also putative recombined RNAs, which may contain the 5' exon from pBD5A (for example, 141 nucleotides, and bands above 464 nucleotides; see Fig. 4 legend). These results show that significant levels of reverse-spliced products can be generated, even at high GTP and low Mg^{2+} concentrations comparable to those *in vivo*. The different results obtained with the *Neurospora* and *Tetrahymena* introns could reflect different equilibrium constants for different intron/exon combinations, or the ability of the CYT-18 protein to shift the equilibrium by binding more strongly to the precursor RNA than to the excised intron.

In conclusion, our results show that reverse splicing of a group I intron can be promoted by a host protein, mitochondrial tyrosyl-tRNA synthetase, required for splicing the intron and that this reaction can compete with forward splicing under physiologically relevant conditions. By speeding the reactions, the CYT-18 protein or other protein catalysts may permit higher levels of reverse-spliced products to be generated at relatively low RNA concentrations *in vivo*, before the intron RNAs are consumed by degradative pathways. The CYT-18 protein functions in splicing of a number of different group I introns¹⁶. Thus, it could promote integration of a variety of introns into different RNAs having short target sequences recognized by exon-binding sites within the introns. As reverse transcriptase activity has been demonstrated in *Neurospora* mitochondria¹⁷, reverse splicing could be a significant pathway for transposition of the

introns to other mitochondrial DNA locations. More generally, other proteins, such as the nuclear-encoded splicing factors identified in *Saccharomyces cerevisiae* mitochondria and maturases encoded by the introns themselves⁸, could also contribute to intron mobility by promoting reverse splicing. From another perspective, our results also demonstrate a novel activity of a tyrosyl-tRNA synthetase. □

Received 5 August; accepted 12 September 1991.

1. Lambowitz, A. M. *Cell* **56**, 323–326 (1989).
2. Cech, T. R. *A. Rev. Biochem.* **59**, 543–568 (1990).
3. Woodson, S. A. & Cech, T. R. *Cell* **57**, 335–345 (1989).
4. Augustin, S., Müller, M. W. & Schweyen, R. J. *Nature* **343**, 383–386 (1990).
5. Mörl, M. & Schmelzer, C. *Cell* **60**, 629–636 (1990).
6. Cech, T. R. *Int. Rev. Cytol.* **3**, 3–22 (1985).
7. Sharp, P. A. *Cell* **42**, 397–400 (1985).
8. Lambowitz, A. M. & Perlman, P. S. *Trends Biochem. Sci.* **15**, 440–444 (1990).
9. Guo, Q., Akins, R. A., Garriga, G. & Lambowitz, A. M. *J. Biol. Chem.* **266**, 1809–1819 (1991).
10. Akins, R. A. & Lambowitz, A. M. *Cell* **50**, 331–345 (1987).
11. Garriga, G. & Lambowitz, A. M. *Cell* **46**, 669–680 (1986).
12. Majumder, A. L. *et al. Molec. cell. Biol.* **9**, 2089–2104 (1989).
13. Kittle, J. D., Jr, Mohr, G., Ganielos, J. A., Wang, H. & Lambowitz, A. M. *Genes Dev.* **5**, 1009–1021 (1991).
14. Bartel, D. P., Doudna, J. A., Usman, N. & Szostak, J. W. *Molec. cell. Biol.* **11**, 3390–3394 (1991).
15. Woodson, S. A. & Cech, T. R. *Biochemistry* **30**, 2042–2050 (1991).
16. Collins, R. A. & Lambowitz, A. M. *J. molec. Biol.* **184**, 413–428 (1985).
17. Kuiper, M. T. R. & Lambowitz, A. M. *Cell* **55**, 693–704 (1988).

ACKNOWLEDGEMENTS. We thank D. Herschlag for comments on the manuscript and J. A. Ganielos for technical assistance. This work was supported by a post-doctoral fellowship from the Deutsche Forschungsgemeinschaft to G.M. and a grant from NIH to A.M.L.

Base-pairing shift in the major groove of (CA)_n tracts by B-DNA crystal structures

Youri Timsit, Eric Vilbois & Dino Moras

Laboratoire de Cristallographie Biologique,
Institut de Biologie Moléculaire et Cellulaire, 15 rue Descartes,
6700 Strasbourg, France

THE crystal packing of the B-DNA dodecamer d(ACCG-GCGCCACA)·d(TGTGGCGCCGGT) is characterized by the reciprocal fit of double helices with specific base-backbone interactions in the major groove. Cooling the crystals below -10°C stabilizes a new conformational state with a long-range sequence-dependent one-step shift in the major-groove base pairing. The tilt of the bases leads to the disruption of the Watson-Crick pairing in the major groove and to the formation of interactions with the 5' neighbour of their complement. This alteration propagates along the helical axis over more than half a turn. As a result, the molecular structure is normal when seen from the minor groove side and mismatched in the major groove. Comparison with a parent isomorphous dodecamer structure corresponding to the codon 10–13 of the *c-Ha-ras* proto-oncogene shows that this new structural feature is sequence dependent and clearly favoured by (CA)_n tracts. As (CA)_n tracts of DNA are involved both in recombination and in transcription, this new recognition pattern should be considered in the analysis of the various processes involving the reading of the genetic information.

The X-ray structure of a nonpalindromic DNA dodecamer containing the *NarI* restriction site GGCGCC, a 'hot spot' for AAF mutagenesis¹ has been described previously². The duplex crystallizes in the B form of DNA but with major deviations from the canonical structure, as the packing induces both the opening and the unstacking of G·C base pairs. In the crystal, specific H-bonds are formed between the phosphate anionic oxygen atoms of the backbone and the amino group of alternate

cytosine bases C3–C21 and C6–C18. These bonds maintain the two helices locked together by a reciprocal major groove/backbone interaction. These four cytosines constitute a packing driving box which orients the crystallization of the duplexes in the R3 space group (Fig. 1a). The design of new sequences where this box is conserved allowed us to obtain isomorphous crystals of several dodecamer duplexes. While searching to improve the quality of the diffraction data, it was observed that the lifetime of the crystals in the X-ray beam increased and the resolution limit was extended from 2.9 to 2.5 Å, when crystals were cooled below -10°C during the data collection (Table 1). Here the -15°C structure of the tet₁₂ dodecamer is compared with that of ras₁₂, d(GGCGCCGGCGGT)·d(ACCGCCGGCGCC) which corresponds to the region 648–659 of the wild type *c-Ha-ras1* gene. This fragment contains the *NarI* sequence which is adjacent to the 5' end of codon 12 where single point mutations are found to activate the *ras* proto-oncogene in many tumours³ (Fig. 1a). The -15°C structures were refined to 15.3% for tet₁₂ and 16.1% for ras₁₂.

The tet₁₂ and ras₁₂ dodecamers present the same local packing-induced alteration, but there are major differences between the two structures in the helix halves with different sequences not involved in the packing contacts. The map of the tet₁₂ dodecamer (Fig. 1b) unambiguously shows a concerted shift in the base pairing from G7 to A12: viewed from the major groove, the bases are not paired with their Watson-Crick complements but with their direct 5' neighbour. The canonical donor-acceptor interactions in the major groove are disrupted to form new ones with the 5' neighbour of the opposite strand. The schematic view of Fig. 2 indicates that the disruption of the G7·C18 base pair is triggered by the attraction of C18 out of the major groove through the formation of an intermolecular N4(C18)–O2P(C20) H bond. The high value of the propeller twist (-32°) at this step induces a close approach between the O6 atoms of the G7 and G17 bases. Electrostatic repulsions are compensated by solvent bridges. On each strand, the bases are tilted in the 3' direction in a domino-like motion with the consequence that the distances N4(C8)–O6(G17), N4(C9)–O6(G16), N6(A10)–O4(T15), N4(C11)–O6(G14) and N6(A12)–O4(T13) of the complementary base pairs are increased well above the normal distance

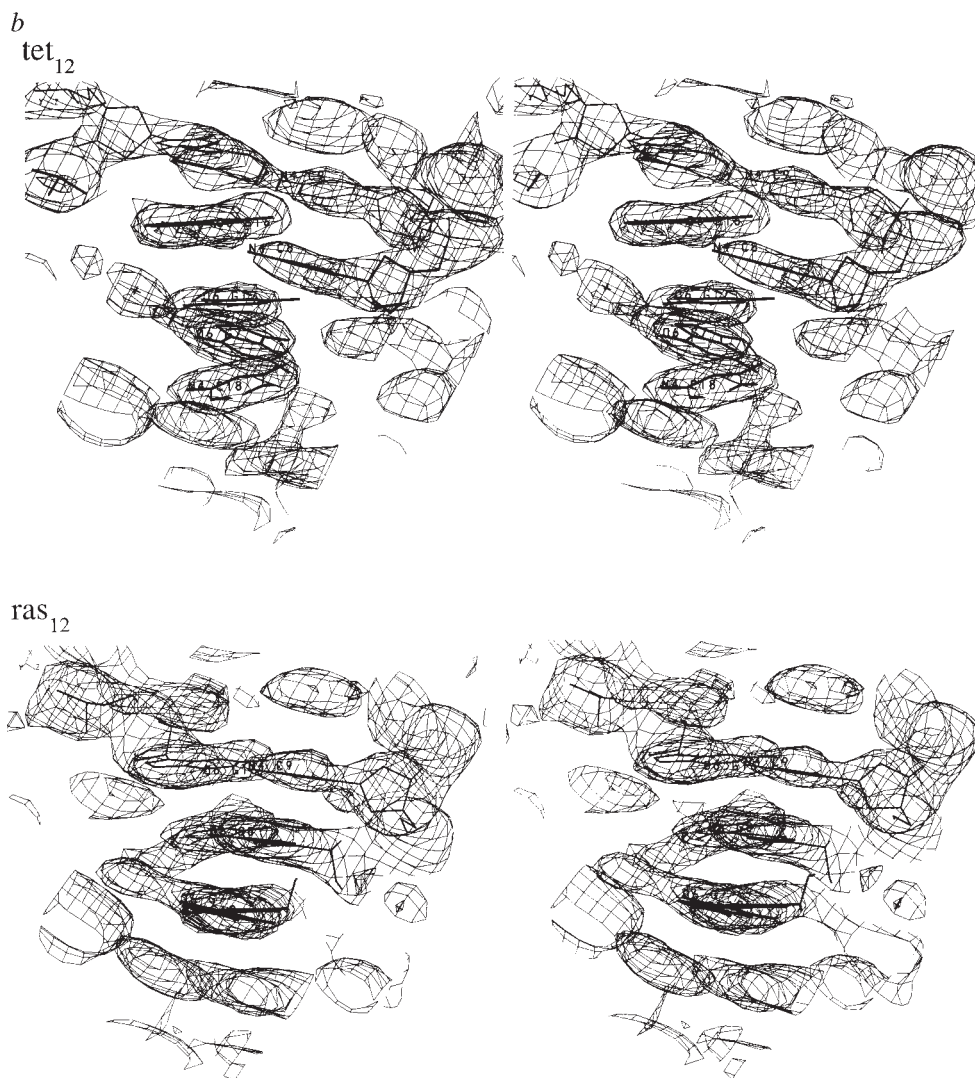
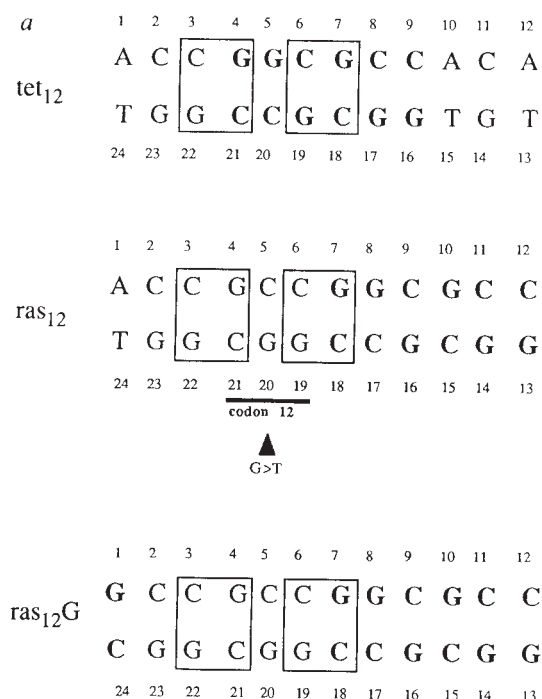


FIG. 1 **a**, Sequences of the tet₁₂, ras₁₂ and ras₁₂ G dodecamers. The tet₁₂ sequence corresponds to the most mutagenic *NarI* restriction site of tetracycline gene in pBR322 plasmid². The ras₁₂ sequence corresponds to residues 648–658 of the *c-Ha-ras* proto-oncogene. The ras₁₂ G sequence corresponds to the ras₁₂ sequence where the first A-T base pair is replaced by G-C. In both sequences, the alternate cytosines 3/21 and 6/18, involved in the groove-backbone interaction are conserved and constitute a packing driving box which orients the crystallization of the duplexes in the R3 space group. In the ras₁₂ dodecamer the most frequent point mutation which activates the proto-oncogene is indicated by the arrow. The *NarI* restriction site is written in bold characters. **b**, Stereo views of 2F_o - F_c maps showing the shifted pairing in the major groove of the G7-C9 regions of the tet₁₂ structure and the normal conformation in the corresponding region of the ras₁₂ dodecamer. The unstable shifted conformation requires both low temperature and spermine to be stabilized, in contrast to the dependence of the diffraction power on the magnesium: oligonucleotide ratio concentrations, common to all other isomorphous structures. This underlines the dynamic aspect of this structure. The shifted paired bases were manually fitted in 3F_o - 2F_c fragment maps calculated without the corresponding residues in the model. Two low-temperature independent data sets of the tet₁₂ dodecamer, provide the same density maps. We used 1,603 and 1,719 unique reflections above the 2σ(F_o) level for tet₁₂ and ras₁₂ structure refinement, respectively. The models with the proper sequence were superimposed on the isomorphous 0 °C parent structure d(ACCGGCGC-CACA)·d(TGTGGCGCCGGT). Atomic coordinates and thermal parameters were first refined with the constrained/restrained least square program CORELS²⁶ using the reflections up to 3 Å. The refinement was then continued with the program NUCLSQ²⁷ with a gradual extension of the resolution up to 2.5 Å without restraint for the base pairing. There were 65 and 58 solvent molecules included, respectively. The final *R* factor for the 1,481 reflections in the 10–2.6 Å range is 16.1% for the tet₁₂ dodecamer. It is 16% for the 1,717 reflections of the ras₁₂ dodecamer.

TABLE 1 Cell parameters and data sets for the isomorphous dodecamers at various temperatures

	Temperature (°C)	Cell parameters (Å)		Cell volume (Å ³)	<i>n</i> ref. >2s(<i>F</i> ₀)	<i>n</i> obs. >2s(<i>F</i> ₀)	<i>R</i> _{sym} (%)	Res. max (Å)
d(ACCGGCCACACA)	0	67.0	48.0	186,594	1,017	2,951	8.9	2.9
d(ACCGGCCACACA)	-10	66.0	46.9	176,621	1,311	2,440	6.5	2.7
d(ACCGGCCACACA)	-15	64.5	47.5	171,375	1,603	3,519	6.2	2.5
d(ACCGCCGGCGCC)	7	65.5	46.5	172,676	1,632	3,662	3.2	2.6
d(ACCGCCGGCGCC)	-15	65.1	47.0	172,712	1,790	3,010	3.5	2.4
d(GCCGCCGGCGCC)	-15	66.3	46.7	172,542	2,503	6,670	5.4	2.3

Data were collected using a Siemens area detector equipped with a rotating anode loaded at 40 kV and 50 mA. The crystals were equilibrated for 30 min in a beam of cooled air. At -15 °C, all crystal forms are isomorphous with the tet₁₂ structure at 0 °C, but a better mosaicity, a longer lifetime of the crystal in X-rays, thinner fibre pattern at 3.4 Å and an extension of the resolution from 2.9 to 2.5 Å are observed. As previously described², tet₁₂ crystals were grown by the vapour diffusion technique, from a solution containing 50 mM cacodylate at pH 6, 1 mM dodecamer duplex, 1.2 mM spermine tetrachloride, 16 mM magnesium acetate and 45% 2,4-methylpentanediol (after equilibration). Both duplexes crystallize under the same conditions, but ras₁₂ and ras₁₂ G crystals can be grown at room temperature and without any spermine.

range for H-bonds, with maximum values (>4.7 Å) at the A-T steps. The conformation is stabilized by new hydrogen interactions which produces the N4(C8)-O6(G16) shifted 'Watson-Crick' and several shifted 'mismatched' major groove pairings, N4(C9)-O4(T15), N6(A10)-O6(G14) and N4(C11)-O4(T15). The shift is communicated to the next stacked molecule through an intermolecular N6(A10)-O4(T24) H-bond at the interhelical step and it continues up to the C3-C21 anchoring point (Fig. 3a).

The isomorphous ras₁₂ dodecamer structure does not have a similar shifted pairing. As in the tet₁₂ structure, the C3-C21 and C6-C18 cytosines are involved in the groove-backbone interaction but the unstacking is less pronounced and the C-G base pairs are not disrupted. Figure 3b shows that the base orientations are very different and the base planarity is less affected. In this case, the formation of bifurcated H-bonds N4(C17)-O6(G7) and N4(C11)-O6(G13) is not accompanied by the disruption of the normal pairing. A detailed comparison of the structures will be described elsewhere.

Several crystal structures of various mismatched oligonucleotides⁴⁻⁸ and the recent observation of the three centred H bonds⁹⁻¹⁰ have shown that the DNA helical structure can accommodate various H-bonding schemes for the base pairing. The shift of the base pairing in the major groove is a new conformational feature of the B-DNA molecule. This unusual structure occurs in an intermolecular contact-free region. It is not observed in the isomorphous dodecamer with a different sequence, suggesting that the shifted structure is associated with the -CCACA- sequence of the tet₁₂ dodecamer. If the conformation can be easily generalized for (CA)_{*n*}/(GT)_{*n*} tracts, it could be asked whether a random set of adenines and cytosines on the same strand can adopt this conformation. Both bases can provide a 'major groove donor strand' for the base pairing. This observation agrees with previous experiments which have suggested unusual structures in (CA)_{*n*} sequences¹¹⁻¹⁵. In the structure of d(GTGTACAC) which crystallizes in the A form, three-centred H-bonds were observed between O6 of guanines and N6 of adenines¹⁶. Moreover, the CA step stacking patterns in B DNA structures are highly variable: in the monoclinic dodecamer d(CCAACGTTGG)¹⁷ and its isomorphous structures they have an unusual 48° high-twist value but not in the orthorhombic dodecamer family.

The shifted structure associated with (CA)_{*n*} tracts combines two properties: the strongly modified structure of the major

groove can constitute a new recognition element for DNA binding proteins, and the slippage of the pairing induces a transient linear modification of the complementarity code which could interfere with the processes associated with the decoding of the genetic information. This could explain their important biological role¹⁸. It was suggested that the common unusual structural feature of CAC triplets can be used for gene regulation and recombination¹². CAC/GTG sequences are required in a number of transcription regulation regions^{11,19}. (CA)_{*n*} tracts are

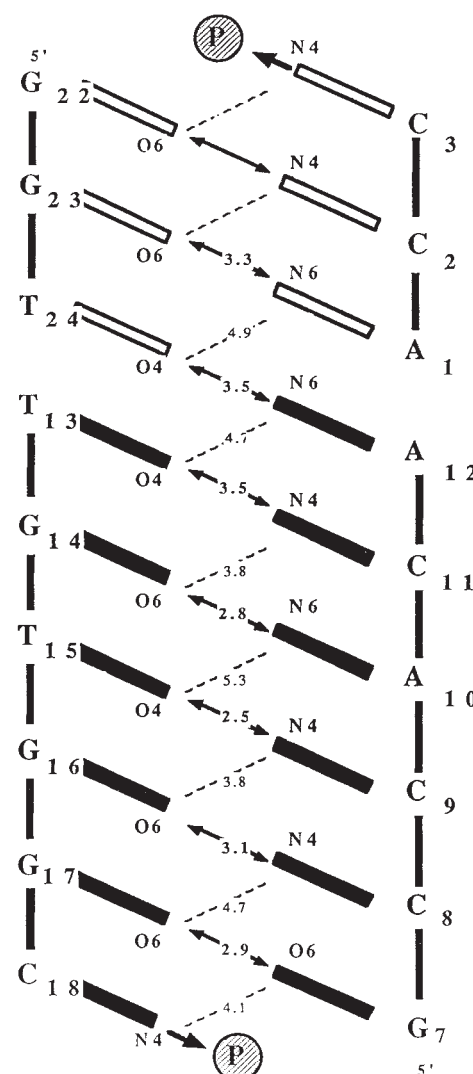


FIG. 2 Schematic view of the major groove of the tet dodecamer in the shifted pairing region. Distances between the donor acceptor base groups of the shifted base pairs are indicated by dark arrows, whereas homologous Watson-Crick H-bond distances are indicated by dotted lines. In the corresponding region of the ras₁₂ dodecamer, the Watson-Crick base pairs remain unaltered.

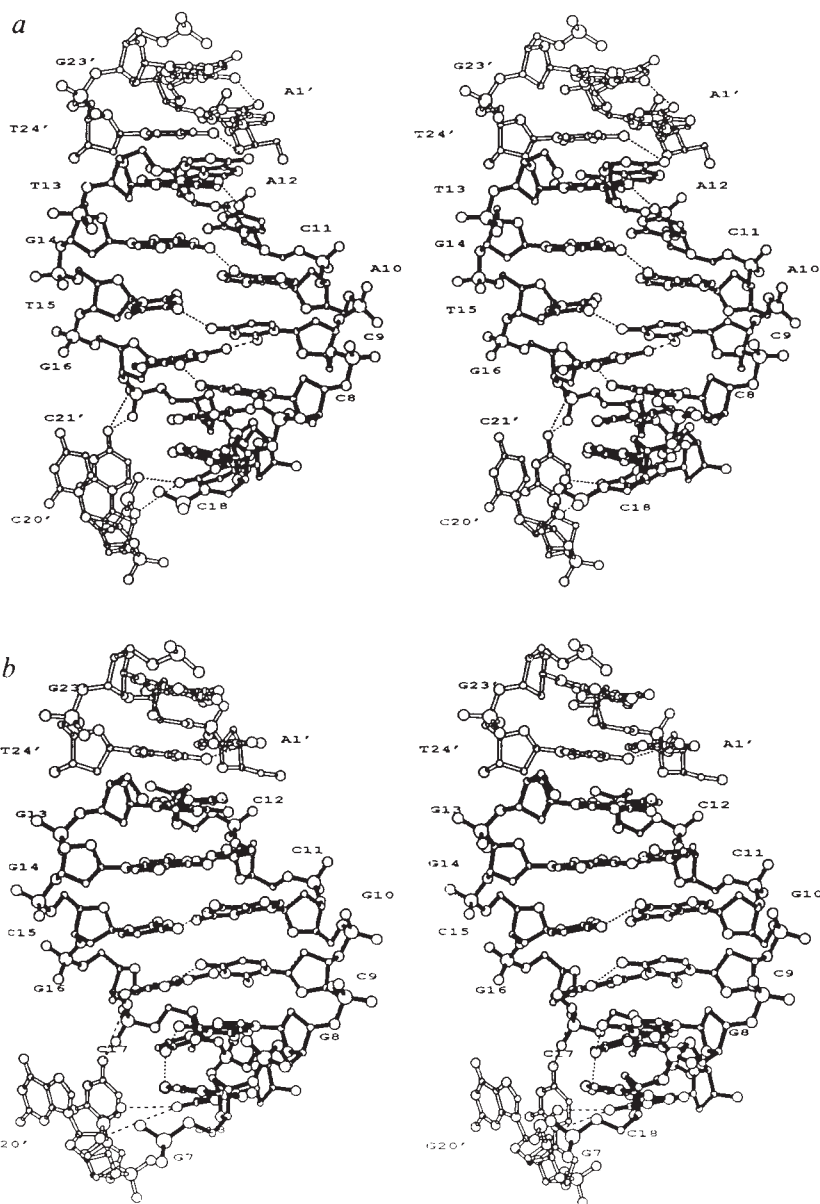


FIG. 3 *a*, Stereo view of the G7-A12 part of the tet₁₂ dodecamer. Shifted H bonds are indicated with dotted lines. The long-range shifted pairing in the tet₁₂ dodecamer is the first visualization of the propagation along the helical axis of a locally altered conformation. More than half a turn of helix is involved in the shift which is also extended to symmetry related duplexes. This accounts for a total of nine shifted base pairs located between the two anchoring points of the crystal. *b*, Stereo view of the same region in the ras₁₂ structure.

involved in gene conversion, insertion events and in the enhancement of recombination²⁰. They may also be implicated in gene duplication mechanisms²¹. This shifted structure could be a starting point to understand the mechanism of recombination such as the unequal crossing over and the imprecise *V-D-J* joining responsible for the immunoglobulin diversity²². It was proposed that this capacity to tolerate misalignments could be required at the hot spot sites of recombination²³.

The transient misalignment of the base pairing could constitute a direct molecular model for the intermediate state of frame shift mutagenesis²⁴ or for the dislocation mechanism proposed by Kunkel²⁵ to explain some substitution mutagenesis events, without the need for looped-out bases. Such sequence-dependent transient misalignments could also be involved in the stabilization of mismatches and could explain the influence of the neighbouring bases in the mutagenesis. □

Received 9 July; accepted 16 September.

1. Fuchs, R. P. P., Schwartz, N. & Daune, M. *Nature* **294**, 657-659 (1981).
2. Timsit, Y., Westhof, E., Fuchs, R. P. P. & Moras, D. *Nature* **341**, 459-462 (1989).
3. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. *Nature* **300**, 149-152 (1982).
4. Hunter, W., Brown, T., Anand, N. N. & Kennard, O. *Nature* **320**, 552-555 (1986).
5. Hunter, W., Brown, T. & Kennard, O. *Nucleic Acids Res.* **15**, 6589-6606 (1987).
6. Kennard, O. *J. biomolec. Struct. Dyn.* **3**, 205-226 (1985).
7. Webster, G. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6693-6697 (1990).
8. Privé, G. et al. *Science* **238**, 498-504 (1987).
9. Coll, M., Frederick, C. A., Wang, A. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8385-8389 (1987).
10. Nelson, H., Finch, J., Luisi, B. & Klug, A. *Nature* **330**, 221-226 (1987).
11. Lu, P., Cheung, S. & Arndt, K. *J. biomolec. Struct. Dyn.* **1**, 509-521 (1983).
12. Cheung, S., Arndt, K. & Lu, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3665-3669 (1984).
13. Usdin, K. & Furano, A. *J. biol. Chem.* **264**, 20736-20743 (1989).
14. McNamara, P., Bolshoy, A., Trifonov, E. & Harrington, R. *J. biomolec. Struct. Dyn.* **8**, 529-538 (1990).

15. Bolshoy, A., McNamara, P., Harrington, R. & Trifonov, E. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2312-2316 (1991).
16. Jain, S., Zon, G. & Sundaralingam, M. *Biochemistry* **28**, 2360-2364 (1989).
17. Privé, G., Yanagi, K. & Dickerson, R. *J. molec. Biol.* **217**, 177-199 (1991).
18. Rodgers, J. *Nature* **305**, 101-102 (1983).
19. Cowie, A. & Myers, R. *Molec. cell. Biol.* **8**, 3122-3128 (1988).
20. Treco, D. & Arnheim, N. *Molec. cell. Biol.* **6**, 3934-3947 (1986).
21. Shen, S., Slightom, J. & Smithies, O. *Cell* **26**, 191-203 (1981).
22. Tonegawa, S. *Nature* **302**, 575-581 (1983).
23. Jarman, A. & Wells, R. *Trends Genet.* **5**, 367-371 (1989).
24. Streisinger, G. et al. *Cold Spring Harb. Symp. quant. Biol.* **31**, 77-84 (1966).
25. Kunkel, T. & Soni, A. *J. biol. Chem.* **263**, 14784-14789 (1988).
26. Sussman, J. L., Holdbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **A33**, 800-804 (1977).
27. Westhof, E., Dumas, P. & Moras, D. *J. mol. Biol.* **184**, 119-145 (1985).

ACKNOWLEDGEMENTS. We thank A. Hoefft for the synthesis of the oligonucleotides and R. Fuchs and G. Webster for reading the manuscript.